P

600
19.2.77

ACKNOWLEDGMENT

To the following persons who have aided in reaching decisions concerning manuscripts submitted for publication during the past year the Editor extends his thanks and appreciation:

George Barlow
University of Illinois

Joseph Becker
University of Washington

Jozef Cohen
University of Illinois

Charles W. Eriksen
University of Illinois

Harold W. Hake
University of Illinois

Murray Miron
Syracuse University

Garth J. Thomas
University of Illinois

Benton Underwood
Northwestern University

Lloyd G. Humphreys, Editor

CONTENTS OF VOLUME 64

Avant, Lloyd L. Vision in the Ganzfeld ... 246

Becker, Gordon. See DeLamater, John.

Bogartz, Richard S. The Criterion Method: Some Analyses and Remarks ... 1

Cahilly, Glenn. See Karrer, Rathe.

Castellan, N. John, Jr. On the Partitioning of Contingency Tables ... 330

Collins, L. Glenn. Comments on "Need States and Activity Level": A Critique ... 80

Dalal, Sudhir N. See Overall, John E.

DeLamater, John, McClintock, Charles G., and Becker, Gordon. Conceptual Orientations of Contemporary Small Group Theory ... 402

Dukes, William F. $N = 1$ ... 74

Frank, George H. Role of the Family in the Development of Psychopathology ... 191

Gerbrandt, Lauren K. Neural Systems of Response Release and Control ... 113

Gogel, Walter C. Equidistance Tendency and Its Consequences ... 153

Greenberg, Marshall G. A Modification of Thurstone's Law of Comparative Judgment to Accommodate a Judgment Category of "Equal" or "No Difference" ... 108

Hartlage, Lawrence C. Effects of Chlorpromazine on Learning ... 235

Jenkins, Herbert M. Measurement of Stimulus Control during Operant Conditioning ... 365

Kahneman, Daniel. Control of Spurious Association and the Reliability of the Controlled Variable ... 326

Kalbfleisch, J. G. See Sprott, D. A.

Karrer, Rathe, and Cahilly, Glenn. Experimental Attempts to Produce Phenylketonuria in Animals: A Critical Review ... 52

Kaufmann, Harry. Definitions and Methodology in the Study of Aggression ... 351

Lawler, Edward E. See Porter, Lyman W.

Leitenberg, Harold. Is Time-Out from Positive Reinforcement an Aversive Event? A Review of the Experimental Evidence ... 428

Levin, Joseph. Three-Mode Factor Analysis ... 442

Levita, Eric. See Riklan, Manuel.

Lewis, Michael. Psychological Effect of Effort ... 183

Lott, Albert J., and Lott, Bernice E. Group Cohesiveness as Interpersonal Attraction: A Review of Relationships with Antecedent and Consequent Variables ... 259

Lott, Bernice E. See Lott, Albert J.

Mackintosh, N. J. Selective Attention in Animal Discrimination Learning ... 124

Malmo, Robert B. Physiological Gradients and Behavior ... 225

McClintock, Charles G. See DeLamater, John.

Meissner, W. W. Functional and Adaptive Aspects of Cellular Regulatory Mechanisms ... 206

Natsoulas, Thomas. Converging Operations for Perceptual Defense ... 393

Overall, John E., and Dalal, Sudhir N. Design of Experiments to Maximize Power Relative to Cost ... 339

Parton, David A., and Ross, Alan O. Social Reinforcement of Children's Motor Behavior: A Review ... 65

Peterson, Lloyd R. A Note on Repeated Measures in the Study of Short-Term Memory ... 151

Porter, Lyman W., and Lawler, Edward E. Properties of Organization Structure in Relation to Job Attitudes and Job Behavior ... 23

Restle, Frank. Significance of All-or-None Learning ... 313

Riklan, Manuel, and Levita, Eric. Laterality of Subcortical Involvement and Psychological Functions ... 217

Ross, Alan O. See Parton, David A.

SCHAIE, K. WARNER. A General Model for the Study of Developmental Problems 92
SPERLING, SALLY E. Reversal Learning and Resistance to Extinction: A Supplementary Report 310
SPROTT, D. A., AND KALBFLEISCH, J. G. Use of the Likelihood Function in Inference 15
TART, CHARLES T. Toward the Experimental Control of Dreaming: A Review of the Literature 81
TAYLOR, INSUP KIM, AND TAYLOR, MAURICE M. Another Look at Phonetic Symbolism 413
TAYLOR, MAURICE M. See TAYLOR, INSUP KIM.
UTTAL, WILLIAM R. Do Compound Evoked Potentials Reflect Psychological Codes? 377
ZILLER, ROBERT C. Toward a Theory of Open and Closed Groups 164

VOL. 64, NO. 1 JULY 1965

Psychological Bulletin

THE CRITERION METHOD:

SOME ANALYSES AND REMARKS

RICHARD S. BOGARTZ[1]

Institute of Child Behavior and Development, University of Iowa

The distribution of trials to criterion under the null hypothesis of no learning is obtained for 2 broad classes of criteria. Tables are provided which give the maximum permissible number of trials which may be run such that the probability of reaching criterion when learning does not occur remains less than a selected value. These results include Grant's analysis of the "runs" criterion as a special case. Illustrations of a number of problems which arise in the use of the criterion method also are presented to emphasize the often ignored dependence of method upon theory.

In the study of learning and transfer, psychologists frequently employ criteria to determine when to terminate the existing treatment condition. In most instances the selection of a criterion is arbitrary, and although some attention has been paid to a number of problems which arise in the use of criteria (Anderson, 1963; Hilgard, 1938, 1951; Underwood, 1954), there is still little in the way of detailed analysis to facilitate efficient criterion selection. A notable exception is Grant's (1946, 1947) treatment of the "runs" criterion and the problem of chance attainment.

Grant (1947) pointed out that if it is possible for the subject to make a correct response by chance, then given enough trials he will reach even the most severe criterion, although his performance level does not actually depart from chance. Grant's concern was with the class of criteria which require a run of n consecutive correct responses. He published tables giving for various values of n and p (the chance probability of a correct response), the largest number of trials which could be run such that the probability of reaching criterion by chance remained less than some selected value. In the present report this probability that criterion is reached, given that the hypothesis of chance performance is true, will be referred to as the criterion risk (CR).

The problem of the CR is general. Every criterion, for a given value of p, determines a CR for each possible number of trials for which the experiment can be run. That is to say, given a constant chance probability of a correct response on each trial, any criterion defines a probability density function, and therefore a distribution function of chance attainment over trials. Consequently, it would be useful to also have tables comparable to Grant's for other criteria.

Many of the more commonly used criteria are members of one or the other of two classes. The first class (Class I) involves an analysis based on blocked trials. The general case in this class is the criterion of at least K correct responses in each of L consecutive blocks of M trials. For example, the subject reaches criterion if he makes at least 9 correct responses on each of 2 consecutive blocks of 10 trials, $(K, L, M) = (9, 2, 10)$; another instance would be 10 correct responses in 1 block of 10 trials, $(K, L, M) = (10, 1, 10)$. The second class (Class II) involves a running analysis over unblocked trials. The general case is K correct responses in any M consecutive trials; for example, 9 correct responses in any 10 consecutive trials. It can be

---

[1] The author is indebted to Ellen Brewer, who was responsible for the computer work throughout most of this investigation, and to Joan Cantor, who made some helpful suggestions concerning arrangement of the results for the Class I and Class II criteria.

TABLE 1

Maximum Permissible Number of Trials at CRs of .01, .05, and .10 for Criteria of at Least K Correct Responses in $L = 1$ Block of M Trials, $p = \frac{1}{2}$

	$K:M$									
CR	5:5	6:6	7:7	8:8	7:8	9:9	8:9	10:10	9:10	11:11
.01	—[a]	—[a]	7	16	—[a]	45	—[a]	100	—[a]	220
.05	5	18	42	104	8	234	18	520	40	1,155
.10	15	36	91	208	16	477	45	1,070	90	—[b]
	10:11	9:11	12:12	11:12	10:12	13:13	12:13	11:13	14:14	13:14
.01	11	—[a]	492	36	—[a]	1,066	65	—[a]	2,296	140
.05	88	11	—[b]	192	24	—[b]	377	52	—[b]	784
.10	187	33	—[b]	396	60	—[b]	793	117	—[b]	1,610
	12:14	14:15	13:15	12:15	15:16	14:16	13:16	16:17	15:17	14:17
.01	14	300	30	—[a]	608	64	—[a]	1,241	136	17
.05	98	1,575	195	30	3,152	384	64	—[b]	731	136
.10	224	—[b]	420	75	—[b]	800	144	—[b]	1,513	272
	17:18	16:18	15:18	14:18	17:19	16:19	15:19	18:20	17:20	16:20
.01	2,484	270	36	—[a]	513	76	19	980	140	20
.05	—[b]	1,692	234	54	2,660	437	95	—[b]	780	160
.10	—[b]	2,880	486	108	—[b]	893	190	—[b]	1,620	340

[a] Criterion risk is exceeded in the first $L \cdot M$ trials.
[b] More than 200 M trials may be run without exceeding the criterion risk.

TABLE 2

Maximum Permissible Number of Trials at CRs of .01, .05, and .10, for Criteria of at Least K Correct Responses in $L = 1$ Block of M Trials, $p = \frac{1}{3}$

	$K:M$									
CR	5:5	6:6	7:7	6:7	8:8	7:8	9:9	8:9	7:9	9:10
.01	10	42	147	7	520	24	1,773	90	9	280
.05	60	222	784	49	—[b]	152	—[b]	477	54	1,440
.10	125	456	—[b]	105	—[b]	320	—[b]	981	108	—[b]
	8:10	7:10	10:11	9:11	8:11	10:12	9:12	11:13	10:13	9:13
.01	20	—[a]	847	77	11	216	24	611	78	13
.05	150	20	—[b]	407	55	1,128	156	—[b]	403	65
.10	300	50	—[b]	836	121	2,316	324	—[b]	819	143
	12:14	11:14	10:14	12:15	11:15	13:16	12:16	13:17	14:18	15:19
.01	1,708	196	28	525	75	1,376	192	493	1,242	3,135
.05	—[b]	1,036	168	2,685	420	—[b]	1,024	2,550	—[b]	—[b]
.10	—[b]	2,128	364	—[b]	870	—[b]	2,112	—[b]	—[b]	—[b]

[a] Criterion risk is exceeded in the first $L \cdot M$ trials.
[b] More than 200 M trials may be run without exceeding the criterion risk.

TABLE 3

Maximum Permissible Number of Trials at CRs of .01, .05, and .10, for Criteria of at Least K Correct Responses in $L = 1$ Block of M Trials, $p = \frac{1}{4}$

CR	K:M									
	5:5	6:6	5:6	7:7	6:7	7:8	6:8	8:9	7:9	8:10
.01	50	246	12	1,148	49	208	16	846	63	240
.05	260	—[b]	66	—[b]	266	1,072	96	—[b]	342	1,230
.10	535	—[b]	132	—[b]	546	—[b]	192	—[b]	702	—[b]
	7:10	9:11	8:11	7:11	9:12	8:12	10:13	9:13	10:14	11:15
.01	20	869	88	11	300	36	1,027	130	406	1,305
.05	140	—[b]	473	66	1,560	216	—[b]	663	2,100	—[b]
.10	300	—[b]	968	143	—[b]	444	—[b]	1,378	—[b]	—[b]

[b] More than 200 M trials may be run without exceeding the criterion risk.

seen that Grant's runs criterion is a member of both these classes. In Class I it is of the form (1, n, 1), and in Class II it is of the form n correct in any n consecutive trials.

Appendix A gives the derivation required for determining the largest permissible number of trials which may be run at a given value of the CR for a Class I criterion, assuming a chance probability of a correct response equal to p. The corresponding derivation for Class II is presented in Appendix B.

Class I Criteria

To obtain results comparable to those of Grant (1947) for the Class I criteria, a program was written for the IBM 7090 which generated the CR for each of the first 200 blocks of trials for various values of K, L, M, and p. The M ranged from 5 to 20, K from $M-4$ to M, and L from 1 to 3. The values of p were $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$. The program also generated the largest number of trials which may be run at various CRs for these criteria. Some of these results are presented in Tables 1–7.

TABLE 4

Maximum Permissible Number of Trials at CRs of .01, .05, and .10, for Criteria of at Least K Correct Responses in $L = 2$ Consecutive Blocks of M Trials, $p = \frac{1}{2}$

CR	K:M									
	5:5	4:5	6:6	5:6	7:7	6:7	7:8	6:8	8:9	7:9
.01	55	—[a]	252	—[a]	1,162	21	72	—[a]	243	18
.05	275	10	—[b]	30	—[b]	98	344	24	1,233	63
.10	560	20	—[b]	60	—[b]	203	712	48	—[b]	126
	9:10	8:10	7:10	9:11	8:11	10:12	9:12	8:12	11:13	10:13
.01	890	40	—[a]	110	—[a]	336	24	—[a]	1,053	65
.05	—[b]	180	20	550	55	1,692	132	24	—[b]	338
.10	—[b]	370	40	1,122	99	—[b]	264	48	—[b]	676
	9:13	11:14	10:14	12:15	11:15	13:16	12:16	13:17	14:18	15:19
.01	—[a]	182	28	510	45	1,440	128	306	774	2,090
.05	52	910	98	2,535	240	—[b]	576	1,496	—[b]	—[b]
.10	91	1,848	210	—[b]	480	—[b]	1,184	3,060	—[b]	—[b]

[a] Criterion risk is exceeded in the first $L \cdot M$ trials.

[b] More than 200 M trials may be run without exceeding the criterion risk.

TABLE 5

Maximum Permissible Number of Trials at CRs of .01, .05, and .10 for Selected Criteria of at Least K Correct Responses in $L = 2$ Consecutive Blocks of M Trials, $p = \frac{1}{3}$

CR	$K:M$								
	4:5	5:6	5:7	6:8	5:8	7:9	6:9	7:10	6:10
.01	30	198	42	216	16	1,332	54	270	20
.05	135	990	189	1,088	64	—^{b}	270	1,360	100
.10	270	—^{b}	378	—^{b}	120	—^{b}	549	—^{b}	200

[b] More than 200 M trials may be run without exceeding the criterion risk.

TABLE 6

Maximum Permissible Number of Trials at CRs of .01, .05, and .10, for Selected Criteria of at Least K Correct Responses in $L = 2$ Consecutive Blocks of M Trials, $p = \frac{1}{3}$

CR	$K:M$									
	4:5	4:6	3:6	5:7	4:7	5:8	4:8	6:9	5:9	6:10
.01	210	48	24	434	21	112	—^{a}	918	45	270
.05	—^{b}	228	84	—^{b}	77	568	40	—^{b}	207	1,350
.10	—^{b}	468	162	—^{b}	161	1,168	72	—^{b}	414	—^{b}

[b] More than 200 M trials may be run without exceeding the criterion risk.

Thus, Table 1 shows that with a criterion of 8 out of 8 in one block, when $p = \frac{1}{2}$, no more than 104 trials may be run if the CR is to remain less than or equal to .05. It also indicates that with a criterion of 7 out of 8 in one block the CR exceeds .01 in the first block, but that with a criterion of 11 out of 11 in one block, more than 2,200 trials may be run without exceeding a CR of .10.

It will be noted that certain of the more severe criteria are omitted from Table 1 for the larger values of M. For example, when $M=15$, values are given only for $K=12$, 13, and 14, but not for $K = 15$. These values are omitted because the CR remains less than .01 even when the experiment is run for 200 blocks of M trials, so that the entries would be Footnote b for .01, .05, and .10. In general, throughout Tables 1–6, when for a given value of M, the largest value of K which is given is less than M, all larger values of K are described by Footnote b for the three tabled values of the CR.

For purposes of exposition, let a Class I criterion be referred to as a "safe" criterion if, given a value of p, that criterion may be used for 200 blocks of trials without the CR exceeding .01. The term "safe" should not be interpreted as "necessarily desirable" since in practice it will often be advisable to choose criteria which are not safe in the sense in which the term is used here.

When $L = 3$, most of the criteria for which values have been obtained are safe. In Table 7 the smallest values of M are given for which the indicated values of K (i.e., M, $M - 1$, . . . , $M - 4$) constitute safe criteria. Thus, for example, a criterion of $M - 2$ out of M on three consecutive blocks will be safe if M is 11 or greater when $p = \frac{1}{2}$, 8 or greater when $p = \frac{1}{3}$, and 7 or greater when $p = \frac{1}{4}$.

TABLE 7

Summary Table of Criteria which are "Safe" for $L = 3$

P	K				
	M	$M - 1$	$M - 2$	$M - 3$	$M - 4$
$\frac{1}{2}$	5	8	11	14	17
$\frac{1}{3}$	5	6	8	10	12
$\frac{1}{4}$	5	5	7	8	10

Class II Criteria

Comparable results for Class II criteria are presented in Tables 8–11. For values of M greater than 10 an approximation method was employed (see Appendix B). The results given in Table 8 permit comparison of approximate with exact values. The exact values of the maximum permissible number of trials which may be run at values of $K = M - 2$ to M, $M = 5$ to 10, for $p = \frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$, are shown together with the nonzero differences (in parentheses) between the approximate and the exact values. A positive difference indicates an overestimate of the exact value; a negative difference, an underestimate. Thus, for example, the exact value for the criterion 8 out of 9, when $p = \frac{1}{3}$ and the CR is .05, is 107 trials; and the negative difference of 6 indicates that the approximate value is 101.

The approximation is quite good for $K = M$; however, this case is of little interest so far as approximation is concerned since it is the Grant runs criterion for which exact values already have been tabled. For K less than M the approximation appears adequate for most purposes. Within the typical range of trials employed in psychological experiments, the approximation remains close to the exact value. In general, the approximate value departs increasingly from the exact value as the exact value increases (i.e., as K, M, or the CR increases, and as p decreases). It appears that for K less than M the approximation is conservative in that it underestimates the exact value; however, this remains to be proved for the general case.

In Tables 9, 10, and 11, the approximate values of the maximum permissible number of

TABLE 8

Maximum Permissible Number of Trials at CRs of .01, .05, and .10, for Criteria of K Correct Responses in M Consecutive Trials, $p = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}$

p	K	CR	M 5	6	7	8	9	10
$\frac{1}{2}$	M	.01	—[a]	—[a]	7	11	17	28
		.05	6	10	17 (1)	31 (1)	59	112 (1)
		.10	9	16 (1)	31	59	113 (1)	222 (1)
	$M - 1$	.01	—[a]	—[a]	—[a]	7	8	9
		.05	—[a]	5	6	9	13	19
		.10	4	5	8	13	20 (−1)	32 (−1)
	$M - 2$	.01	—[a]	—[a]	—[a]	—[a]	7	8
		.05	—[a]	—[a]	5	6	8	9
		.10	—[a]	—[a]	5	7	9	12
$\frac{1}{3}$	M	.01	7	15	38	105	304	898
		.05	21 (1)	60	173	510 (1)	1,521	—[b]
		.10	41	119	350 (1)	1,042 (1)	—[b]	—[b]
	$M - 1$	.01	—[a]	5	7	13	26 (−1)	60 (−3)
		.05	5	9	19	44 (−2)	107 (−6)	274 (−14)
		.10	7	15	35 (−2)	84 (−4)	213 (−12)	555 (−29)
	$M - 2$	.01	—[a]	—[a]	5	7	9	14
		.05	3	5	7	11	21 (−1)	42 (−3)
		.10	3	5	10	18 (−1)	37 (−3)	80 (−8)
$\frac{1}{4}$	M	.01	17	59	225	884	—[b]	—[b]
		.05	73	284	1,125 (1)	—[b]	—[b]	—[b]
		.10	146 (1)	579 (1)	2,306 (1)	—[b]	—[b]	—[b]
	$M - 1$	.01	4	8 (−1)	18 (−1)	50 (−2)	159 (−6)	548 (−20)
		.05	9	23 (−1)	70 (−3)	229 (−10)	786 (−32)	2,766 (−102)
		.10	15	44 (−2)	139 (−6)	465 (−21)	1,607 (−66)	—[b]
	$M - 2$	.01	—[a]	4	6	10	22 (−1)	55 (−4)
		.05	3	6	13	31 (−2)	84 (−6)	248 (−19)
		.10	4	10	23 (−2)	59 (−5)	167 (−14)	502 (−40)

Note.—Numbers in parentheses are nonzero differences obtained by subtracting exact value from value obtained with approximation method.

[a] Criterion risk exceeded on Trial K.

[b] Criterion risk not exceeded by Trial 3,000.

TABLE 9

Approximate Maximum Permissible Number of Trials at CRs of .01, .05, and .10, for Criteria of K Correct Responses in M Consecutive Trials, $p = \frac{1}{2}$

K	CR	M 11	12	13	14	15	16	17	18	19	20
$M-1$	.01	12	17	24	37	59	101	179	325	602	1,126
	.05	29	47	80	140	252	461	855	1,597		
	.10	51	88	154	277	505	934	1,740			
$M-2$	.01	9	11	12	15	19	25	36	54	85	140
	.05	12	16	22	32	49	77	126	213	368	646
	.10	17	24	36	55	88	144	244	421	738	1,309
$M-3$	.01	8	9	11	12	14	15	18	22	29	38
	.05	9	11	13	16	21	27	38	54	82	128
	.10	10	13	17	22	30	43	63	96	151	245
$M-4$	.01	7	8	9	11	12	14	15	17	19	22
	.05	8	9	11	13	14	17	20	25	33	44
	.10	9	10	12	14	17	22	28	37	51	73

Note.—Empty cells denote approximate values greater than 3,000.

trials for various values of K and M are given, for $p = \frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$, respectively. Approximate values were not obtained if they exceeded 3,000. The only value of $M - 1$ out of M which did not exceed 3,000 when $p = \frac{1}{4}$ was 1,882 for 10 out of 11.

REMARKS

Some remarks may facilitate the use of these results. Other more general remarks are also included which may give the investigator reason to reconsider the desirability of using the criterion method.

Running Analysis. When the same values of K and M are used in both the running analysis and the blocked trial method ($L = 1$), the former always entails a greater criterion risk than the latter (and increasingly so as L increases). The direction of this difference has been known; these results permit evaluation of the magnitude of the difference. It can be seen that in terms of the permissible number of trials, the difference can be sizable.

TABLE 10

Approximate Maximum Permissible Number of Trials at CRs of .01, .05, and .10, for Criteria of K Correct Responses in M Consecutive Trials, $p = \frac{1}{3}$

K	CR	M 11	12	13	14	15	16	17	18	19	20
$M-1$	.01	142	374	1,012	2,781						
	.05	690	1,868								
	.10	1,408									
$M-2$	.01	23	46	102	243	607	1,557				
	.05	84	195	475	1,191						
	.10	164	391	965	2,434						
$M-3$	.01	11	15	24	43	87	192	449	1,089	2,698	
	.05	22	40	80	172	391	923	2,232			
	.10	37	73	153	342	791	1,882				
$M-4$	.01	9	11	13	17	26	43	81	169	374	865
	.05	11	16	25	43	81	165	355	798	1,842	
	.10	16	25	42	77	154	325	715	1,623		

Note.—Empty cells denote approximate values greater than 3,000.

An even more striking result for the running analysis is the difference in the permissible number of trials between the case when $K = M$ and that when K is less than M. For instance, when the chance probability of a correct response is one half, and a CR of .05 is adopted, the permissible number of trials with a criterion of 10 out of 10 is 112; with a criterion of 9 out of 10 it drops to 19!

CR and Error Rates. The CR may be thought of as the Type I error rate in the test of the chance performance hypothesis; that is, the probability of concluding that the subject is performing better than chance when in fact he is not. There is a corresponding Type II error rate, the probability of concluding that the subject's true probability of a correct response is at the chance level when in fact it is greater. This Type II error rate will depend upon the severity of the criterion, the subject's initial probability of a correct response, his learning rate, and the selected value of the CR.

Under certain circumstances the present results will be useful not only in setting the Type I rate, but also in controlling the Type II rate. Practical considerations, such as college students often being available for only 50 minutes, or preschool children being unwilling to remain in most experimental situations longer than 20–30 minutes, may determine the maximum number of trials which can be run in a particular experiment. When this maximum number is known, it will usually be possible to select from among several criteria which have the desired CR that criterion which is least severe, and thus to reduce the Type II error rate. This use of the tables will be increasingly appropriate to the extent that the experimenter is only interested in concluding that the subject is performing better than chance and has little interest in the subject's performance level at attainment of the criterion.

Criteria and Training-Transfer Paradigms. Often the experimenter is not interested in the hypothesis of chance performance, but rather in "equating" subjects whose terminal performance levels would differ if they were all run for the same number of trials. Such terminal differences could be due, for example, to differences in initial performance levels, differences in parameters governing the rate of change in the performance level over trials, or both. The logic of this method of equating subjects can be exemplified with the linear learning model of stimulus sampling theory. Assume that one response, A_1, is reinforced on each trial of the training stage of an experiment. The linear model asserts that for Subject x,

$$p_{x,n} = 1 - (1 - p_{x,1})(1 - \theta_x)^{n-1},$$

where $p_{x,n}$ is the probability of occurrence of A_1 on Trial n, $p_{x,1}$ is the initial probability of that response, and θ_x is the learning rate parameter. Also let $p_{x,t}$ denote the probability of A_1 for Subject x following the terminal trial of the training stage.

TABLE 11

APPROXIMATE MAXIMUM PERMISSIBLE NUMBER OF TRIALS AT CRs OF .01, .05, AND .10, FOR CRITERIA OF K CORRECT RESPONSES IN M CONSECUTIVE TRIALS, $p = \frac{1}{4}$

		M						
K	CR	11	12	13	14	15	16	17
	.01	148	464	1,525				
$M-2$	.05	717	2,328					
	.10	1,463						
	.01	26	60	162	477	1,476		
$M-3$	.05	97	267	783	2,389			
	.10	190	539	1,598				
	.01	11	18	33	73	190	534	1,596
$M-4$	.05	25	52	125	328	918	2,669	
	.10	43	97	245	662	1,872		

Note.—Empty cells denote approximate values greater than 3,000.

If all of the subjects are run for the same number of training trials there will be a distribution of terminal values over the group of subjects. The variance of this distribution will depend upon the variance of the distribution of rate parameters and that of the distribution of initial probabilities. For the purposes of this discussion it will suffice to focus upon the distribution of θ values and assume that all subjects have the same value of p_1. This will correspond to the context of "fast" and "slow" learners in which the use of criteria is most often discussed.

The general problem of equating the subjects is analogous to the problem of reducing the variance of the distribution of p_t values. In terms of the model, the criterion method is an attempt to accomplish this by adjustment of the number of trials a subject receives to his rate parameter so that subjects with high values of θ (fast learners) are run fewer trials than subjects with low values (slow learners). Since $p_{x,t}$ depends upon both θ_x and upon n_x, the number of trials that Subject x is run, the variance in the distribution of p_t values will depend upon the variance of the distribution of θ values, the variance of the distribution of n, and upon the covariance of n with θ. The basic premise of the criterion method is that termination of training at attainment of the criterion produces a covariance of n and θ which more than offsets the variance added to the distribution of p_t by variation in n.

Certain aspects of the criterion method become somewhat more obvious when the method is viewed in these terms. In particular, if all of the subjects have the same learning rate and the criterion method is used, in general they will not all reach criterion on the same trial, just as subjects who are not learning (i.e., performing at the chance level) reach criterion at different trials (see Appendices A and B). (This is because a portion of the variability in the number of trials to criterion is produced by the criterion method itself and has nothing to do with differences in learning rates.) As a consequence, the criterion method introduces variance in the distribution of p_t, due to variance in n, which cannot be offset by covariance of n with θ since all subjects have the same value of θ. Thus the criterion method actually will produce greater variability in the true performance levels than would be obtained if all subjects had been run for the same number of trials.

An obverse implication also follows from the fact that even when homogeneity of learning rates exists, subjects will reach criterion at different trials: If a criterion method is used and it is found that some subjects reach criterion sooner than others, it does not follow that heterogeneity of learning rates exists in the group of subjects, since this same finding would be expected even with homogeneity of learning rates. This implication may have serious consequences so far as interpretation of experiments in which subjects have been separated into fast and slow learners on the basis of their performance as measured by the number of trials to criterion.

Separation of Subjects by the Median Trial to Criterion. The implications of using the median trial to criterion to divide a group of subjects into fast learners and slow learners have not been treated in the psychological literature. Some of these implications are not so obvious as others. Many rest upon the previously mentioned fact that just as there is a distribution of trials to criterion in the case where subjects are not learning, there will also be a distribution when they are learning, and that this holds whether or not the group is homogeneous with respect to learning rates.

In the following discussion it is assumed for illustrative purposes that the learning function for Subject x is $p_{x,n} = 1 - .5(1 - \theta_x)^{n-1}$. Thus each subject has an initial probability of a correct response equal to .5, and therefore the subjects may only differ in their learning rate parameters. It is also assumed that an overall group of subjects comprises two subgroups having the same number of subjects, and that all subjects in the same subgroup have the same learning rate; that is, the subgroups are homogeneous.

Now, given the learning function and a Class I criterion to which the entire group of subjects is to be run (e.g., $[K, L, M] = [1, 6, 1]$; a criterion of six consecutive correct responses), it is possible to compute the probability of reaching criterion on the nth trial, $P(C_n)$. It is then possible to obtain the expected value of the probability of a correct response on the first postcriterion trial, $E(p_t)$, by a summation of products,

$$\sum_{n=1}^{\infty} P(C_n)[1 - .5(1 - \theta)^n].$$

Thus an $E(p_t)$ can be computed for each of the two subgroups comprised by the overall group. These two values will be referred to as the True Slow $E(p_t)$ and the True Fast $E(p_t)$ when computed for the subgroups with the slower and faster learning rates, respectively.

Two other expectations of interest can also be computed. First the median trial to criterion for the overall group must be obtained. If $P_s(C_n)$ is the probability that a subject in the slow subgroup reaches criterion on Trial n, and $P_f(C_n)$ is the corresponding value for a fast subject, then under the assumption of an equal number of subjects in each subgroup, the theoretical median trial to criterion for the overall group will be that value N which

TABLE 12

EFFECT OF USING THE MEDIAN TRIAL TO CRITERION TO SEPARATE "FAST" FROM "SLOW" LEARNERS

Criterion	Method of separation	Learning rate pairs in overall group									
		.01–.01		.01–.03		.01–.05		.01–.07		.01–.09	
		Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
6	True	663	663	663	756	663	806	663	840	663	865
	Median	719	602	749	672	747	720	743	759	739	791
8	True	730	730	730	817	730	860	730	888	730	909
	Median	789	668	797	751	787	803	773	843	765	870
10	True	781	781	781	859	781	895	781	919	781	936
	Median	836	725	832	808	815	861	809	893	803	915
		.03–.03		.03–.05		.03–.07		.03–.09		.03–.11	
		Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
6	True	756	756	756	806	756	840	756	865	756	885
	Median	824	696	840	729	840	761	837	781	834	806
8	True	817	817	817	860	817	888	817	909	817	925
	Median	873	757	883	790	880	826	874	853	869	872
10	True	859	859	859	895	859	919	859	936	859	948
	Median	907	807	914	844	906	869	899	895	895	913

Note.—Decimal points omitted.

minimizes

$$\left| \sum_{n=1}^{N} [P_s(C_n) + P_f(C_n)] - 1 \right|.$$

Let θ_s and θ_f be the two learning rates. Then, given N, the $E(p_t)$ for those subjects in the overall group who reach criterion on or before the median trial, the "fast" subjects, will be

$$\sum_{n=1}^{N} [P_s(C_n)[1 - .5(1 - \theta_s)^n] + P_f(C_n)[1 - .5(1 - \theta_f)^n]].$$

This value, which will be referred to as the Median Fast $E(p_t)$, is the expected probability of a correct response on the first postcriterion trial for a subject labeled "fast" by virtue of the fact that he reached criterion on or before the median trial. The corresponding value for "slow" subjects, the Median Slow $E(p_t)$, is obtained by summing from $N + 1$ to ∞.

In Table 12 the four $E(p_t)$'s, True Slow, True Fast, Median Slow, and Median Fast, are presented for each of 10 different learning rate pairs at each of three different criteria. The criteria are 6, 8, and 10 consecutive correct responses. It will be noted that two of the learning rate pairs, .01–.01 and .03–.03, have identical values for the slow and fast subgroups. These are included to indicate the effect of using the median trial when the entire group is homogeneous. The tabled values illustrate a number of desirable and undesirable features of the criterion method in general and also of the median trial separation procedure. Four of these will be indicated.

First, inspection of the True values indicates that for a given learning rate, the $E(p_t)$ increases as the criterion becomes more severe. For example, with a θ value of .01 the True $E(p_t)$ values are .663, .730, and .781 for the criteria of 6, 8, and 10 consecutive correct responses, respectively. This is in accord with the intuitive expectation that the harder the criterion the higher will be the average performance level when it is achieved (if learning is taking place).

Second, it can be seen that for a given criterion, the True $E(p_t)$ values go up as the learning rate increases. Thus for the criterion of six consecutive correct responses, as the learning rate ranges from .01 to .09, the True Fast $E(p_t)$ ranges from .663 to .865. This exemplifies the point made by a number of authors that the criterion does not exactly equate subjects who are learning at different rates, but rather the faster learner, on the average, will be somewhat above the slower learner on the first postcriterion trial. Although it is the case that the faster the learning rate the higher the $E(p_t)$, it does not follow that the greater the difference in learning rates the greater the disparity in $E(p_t)$'s. For example, it can be seen that for the criterion of six consecutive correct responses, a difference of .06 between the θ values .01 and .07 results in a difference of .177 between the values .663 and .840, but a difference of .08 between the θ values .03 and .11 results in a difference of only .129 between .756 and .885. This is due to a ceiling effect. What can be said is that given a slow group with a certain learning rate, the greater the difference between that given rate and the learning rate for a fast group, the greater will be the disparity in their $E(p_t)$ values.

Third, it can be seen that some of the Median values are inverted. When the learning rates for the two subgroups are identical (both .01 or both .03) the Median Slow $E(p_t)$ is greater than the Median Fast $E(p_t)$. Thus when both subgroups have θ values of .03, the Median Slow and Median Fast values are .824 and .696, .873 and .757, and .907 and .807, for the three criteria, respectively. The implication is obvious: If a group of homogeneous subjects is run to criterion and then divided on the basis of the median into fast and slow learners, the slow learners can be expected to behave as if they learned more than did the fast learners. And, on the average, they did.

It will also be noted that the inversion also can occur when the two subgroups differ in their learning rates. The instances of inversion have been underlined in Table 12. The pattern suggests certain other generalizations but these will not be enumerated here.

The fourth feature of Table 12 indicates what has probably been the intuitive basis for the use of the median trial to criterion as a separating point. This is that as the learning rates for the two subgroups move farther and farther apart, the Median $E(p_t)$ values converge to the True values.

This finding of a possible inversion of the Median Slow and Median Fast $E(p_t)$'s, the occurrence of which depends upon a number of parameters, appears to be of considerable significance with respect to interpretation of experimental results. Suppose, for example, that the median trial to criterion on a training task is used to separate subjects into fast and slow learners, and that on a transfer task the fast learners perform more poorly than do the slow learners. Under the assumption that the terminal level for fast subjects is higher than that for slow subjects, one fairly natural conclusion would be that learning the first task produces interference on the second. The results presented in Table 12 indicate quite clearly that such a conclusion is completely gratuitous. It may actually be the case that learning of the first task is unrelated to performance on the second or that it is even facilitative, since it may actually be the slow learners who on the average have learned more in training. Without information concerning the learning function and the existing rate parameters (which if available would eliminate the need for a criterion) as well as the effects produced by the particular criterion selected, it is not clear how to interpret differences between the fast and slow learners on a transfer task.

"x *Percent Overlearning*" *Procedure.* Another commonly used procedure is to give a subject "overlearning" trials, the number of which depends upon the number of trials the subject takes to reach criterion. Thus in the 100%-overlearning procedure, if the subject reaches criterion on Trial n, he is given another n trials; in 50% overlearning he is given an additional $n/2$ trials; etc.

It has already been observed that the more homogeneous the group of learners the more will the criterion method inflate the variance of their terminal performance levels as compared with an identical group of subjects who are all run for the same number of trials, and that this inflation is exactly the opposite of the goal of using the criterion. Now consider a group of identical subjects who are run to a criterion. Some reach it early; some late. Because they are identical so far as their rate parameters are concerned, the ones who reach criterion early will have lower terminal performance levels than will those who reach it late. If the x%-overlearning method is employed, the subjects who reach criterion early

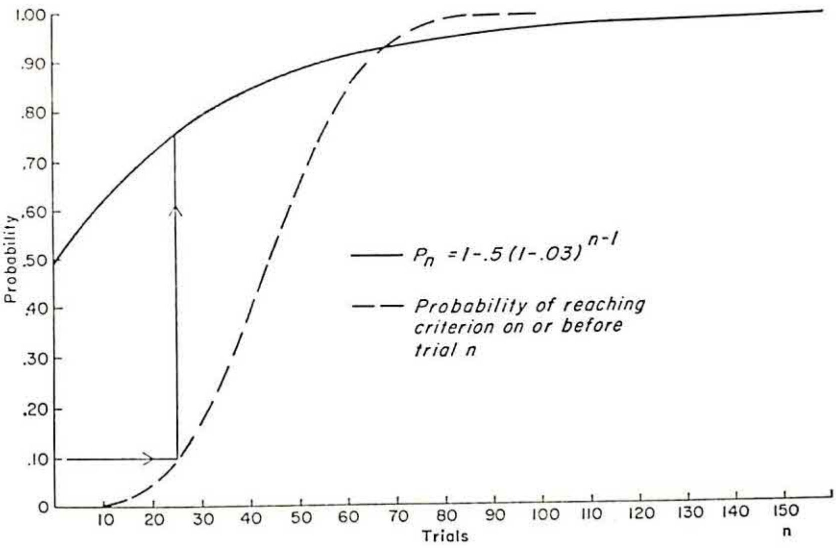


FIG. 1. The probability of reaching a criterion of 10 consecutive correct responses on or before Trial n, given the learning curve $p_n = 1 - .5\ (1 - .03)^{n-1}$.

will receive fewer additional trials than will those who reach it late. Thus the overlearning method amplifies the injustice so far as the total number of trials is concerned. Whether or not it produces an even greater inflation of the variance of terminal performance levels will depend upon the form of the learning function and the particular criterion which is used. When the subjects are not identical the effects will depend upon the distribution of learning rates as well as on the function and the criterion.

There is another problem involved in this procedure. Suppose that some criterion method does equate (at least roughly) the performance levels of a group of subjects learning at different rates. Will the overlearning procedure preserve the equating or will the subjects be spread out again by the addition of different numbers of trials? In general, this will depend upon the form of the function. As an example, let $f(\theta_x, n_x) = 1 - (1 - p_1)(1 - \theta_x)^{n_x - 1}$. It can be shown that if $f(\theta_s, n_s) = f(\theta_f, n_f)$, then $f\left(\theta_s, \left(1 + \frac{x}{100}\right) n_s\right)$ is approximately equal to $f\left(\theta_f, \left(1 + \frac{x}{100}\right) n_f\right)$. This means that in the case of the linear learning model, once subjects learning at different rates have been equated, the $x\%$-overlearning procedure will preserve the equating (except for a negligible amount). But surely this will not be true of learning models in general, and in view of the fact that the form of the learning function is almost always in question, careful consideration should be given to the desirability of using this method.

Inference of the Performance Level from Attainment of the Criterion. The solid curve in Figure 1 is the learning curve $p_n = 1 - .5(1 - .03)^{n-1}$. The dashed curve is a distribution function which gives the probability that a subject having this learning function will have reached a criterion of 10 consecutive correct responses by the nth trial; that is, it is the integral of the density function which gives the probability of reaching criterion on Trial n, given this learning function. If a large group of subjects were all learning according to the given learning function, the dashed curve would give the expected proportion of subjects reaching criterion on or before Trial n.

The two curves can be used to determine confidence intervals. For example, a line drawn parallel to the abscissa from .10 on the ordinate will intersect the dashed curve at the trial upon which 10% of the subjects can be expected to have reached criterion. In Figure 1 this occurs at Trial 25. A line drawn parallel to the ordinate from this point of intersection will intersect the solid curve. The value of p_n at this second point of intersection is then read on the ordinate to be about .76. Ten percent of the subjects will have a value of p_n no greater than .76 on the trial they reach criterion. It may also be said that 90% of the subjects will have a value of p_n greater than or equal to .767 (the

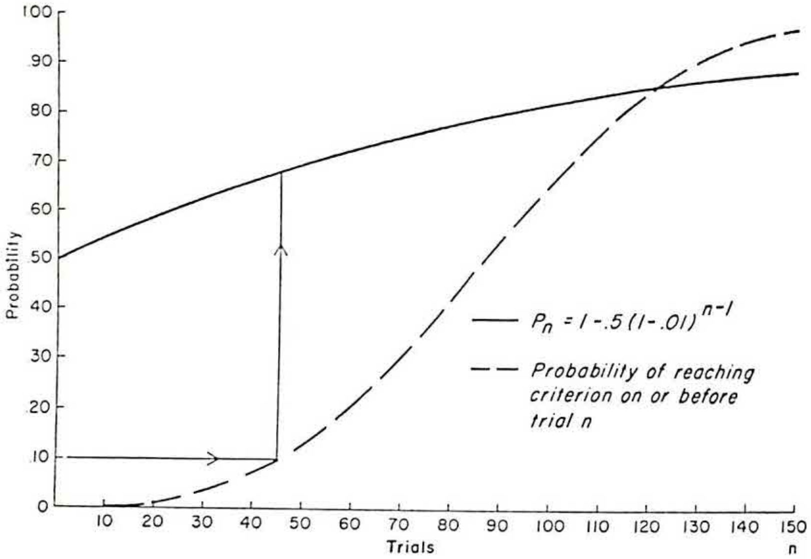


FIG. 2. The probability of reaching a criterion of 10 consecutive correct responses on or before Trial n, given the learning curve $p_n = 1 - .5\ (1 - .01)^{n-1}$.

value of p_n on Trial 26) on their first postcriterion trial.

In Figure 2 the learning curve $p_n = 1 - .5(1 - .01)^{n-1}$ is plotted together with the corresponding distribution function for the same criterion of 10 consecutive correct responses. It can be seen that with this learning curve 90% of the subjects who reach criterion will have a value of p_n greater than or equal to .679. About 50% of the subjects will reach criterion before p_n reaches .80, whereas in Figure 1 only about 18% reach criterion before p_n reaches .80.

These results illustrate the fact that the distribution of terminal performance levels obtained through the use of a criterion will depend not only upon the criterion which is used, but also upon the learning function and the parameters characterizing the group of subjects. Since the learning function and parameters are unknown, it will be difficult to infer the range of performance levels which has been achieved. This is not to suggest that it is impossible to put criteria to such use, but only to indicate that without additional information to be used in conjunction with the selected criterion, it is difficult to support any conjectures as to the terminal performance level resulting from the use of a particular criterion.

Concluding Comments. When the criterion method is used to test the hypothesis that no learning has occurred, there are no implicit assumptions concerning the form of the learning function or the learning rates present in the population of subjects. This is the case because the alternatives are no learning versus learning of any sort. However, when the goal is to equate the terminal levels of a group of subjects, then it appears that additional information is required if one is to choose between using a fixed number of trials and a criterion. It is not sufficient to rest usage of a criterion upon the unlikelihood that all subjects learn at the same rate. It must be shown that the difference in rates makes a difference, and this should not be done by indicating that some subjects reach criterion before others.

There are instances where the criterion method is the most natural way to proceed. For example, in Theios' (1963) study of avoidance conditioning, the model he used assumed subjects are eventually absorbed into a state in which only correct responses occur. He used a criterion of 20 consecutive correct responses. The problem is that this case is an exception. In far too few instances is the model or class of models and the implications for use of the criterion method made explicit.

REFERENCES

ANDERSON, N. H. Comparison of different populations: Resistance to extinction and transfer. *Psychological Review*, 1963, **70**, 162–179.

FELLER, W. *An introduction to probability theory and its applications.* New York: Wiley, 1957.

GRANT, D. A. New statistical criteria for learning and problem solution in experiments involving repeated trials. *Psychological Bulletin*, 1946, 43, 272–282.

GRANT, D. A. Additional tables of the probability of "runs" of correct responses in learning and problem solving. *Psychological Bulletin*, 1947, 44, 276–279.

Hilgard, E. R. A summary and evaluation of alternative procedures for the construction of Vincent curves. *Psychological Review*, 1938, **35**, 282–304.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Theios, J. Simple conditioning as two-stage all-or-none learning. *Psychological Review*, 1963, **70**, 403–417.

Underwood, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychological Bulletin*, 1954, **51**, 276–282.

APPENDIX A

Class I consists of the criteria in which S reaches criterion by making at least K correct responses on each of L consecutive blocks of M trials. The CR for i blocks of M trials will be the probability that criterion is reached on or before the ith block. To obtain this distribution function it will suffice to derive the density function and then cumulate it.

The probability of exactly x correct responses in any block of M trials will of course have the binomial distribution

$$b(x, M, p) = \frac{M!}{x!(M-x)!}\, p^x (1-p)^{M-x}.$$

The probability of at least K correct responses is simply

$$\sum_{x=K}^{M} b(x, M, p)$$

which is denoted by B (suppressing K, M, and p).

In order for the event C_i, attainment of criterion on Block i, to occur, three conditions must obtain:

1. At least K correct responses are made on each of the L consecutive blocks $i - L + 1$ to i (this occurs with probability B^L).

2. At most $K - 1$ correct responses are made in Block $i - L$, otherwise criterion would have been reached on Block $i - 1$ (this occurs with probability $1 - B$).

3. In the first $i - L - 1$ blocks, no more than $L - 1$ consecutive blocks contain more than $K - 1$ correct responses. This simply asserts that criterion has not been reached in the first $i - L - 1$ blocks. (The probability that it has been reached is the distribution function

$$F(i - L - 1) = \sum_{j=1}^{i-L-1} \Pr(C_j);$$

therefore the probability that it has not been reached is $1 - F(i - L - 1)$).

Using the product rule for the probability of the joint occurrence of independent events, we obtain the recursive expression

$$\Pr(C_i) = [1 - F(i - L - 1)](1 - B)B^L$$

for the probability density function.

For $L = 1$ it is easily shown by induction that the probability density function is the geometric distribution, $\Pr(C_i) = B(1 - B)^{i-1}$, with mean B^{-1} and variance $(1 - B)B^{-2}$. The CR is

$$F(i) = \sum_{j=1}^{i} B(1 - B)^{j-1} = 1 - (1 - B)^i.$$

Thus, if we plot, as a function of blocks, the probability that S has reached criterion on or before that block, we obtain for $L = 1$ a negatively accelerated curve with asymptote at unity and with B as the rate parameter—a "learning" curve without learning, so to speak.

The value of B for particular values of K, M, and p is easily obtained with a desk calculator or tables. It is then a simple matter to obtain the largest i such that $F(i)$ is less than or equal to a selected CR.

APPENDIX B

A Class II criterion is of the form K correct responses in any M successive trials. With this type of criterion a running analysis is made over trials 1 to M, 2 to $M + 1$, 3 to $M + 2$, etc., until M successive trials occur which contain K successes. The first such occurrence constitutes attainment of the criterion.

For $n \leq M$ the probability of C_n, attainment on Trial n, is the probability of a success on Trial n and $K - 1$ successes in the $n - 1$ previous trials. We have then

$$\Pr(C_n) = \begin{cases} 0 & n < K \\ \dfrac{(n-1)!}{(K-1)!(n-K)!}\, p^K (1-p)^{n-K} & K \leq n \leq M \end{cases}$$

For $n > M$ the problem is somewhat more complicated. It can be shown, however, that for $n > M$, under the hypothesis of chance performance on each trial, a Class II criterion is a Markov chain with one absorbing state (see Feller, 1957, Ch. XV, for a discussion of Markov chains). Three steps are required to show this: a set of possible states of the chain must be defined, an initial (Trial M) probability distribution for these states must be specified, and a matrix of transition probabilities p_{jk} must be specified. This matrix which gives for all j and k the probability of the system being in State k on the next trial, given that it is in State j on the present trial, must be independent of the trial number.

To define the states of the chain we first consider the 2^M different sequences of zeros and ones (corresponding to errors and correct responses), 111 . . . 11, 111 . . . 10, . . . , 000 . . . 01, 000 . . . 00, which are each of length M. Setting aside the question of criterial performance, we may observe that at each step in the running analysis of a response protocol we have in hand one of these 2^M sequences, and that given that sequence, only two sequences are possible at the next step. For example, if M is 5 and we observe the sequence 00110 at Step n, then at Step $n + 1$ we shall observe either 01101 or 01100. The initial digit on the left is dropped (we move forward one trial), and either a 1 or a 0 is added on the right, depending upon the outcome of the next trial. We observe 01101 if a correct response occurs or 01100 if an error occurs.

Thus if there were no interest in a criterion, each of the 2^M sequences could constitute a state in a Markov chain. Each row of the transition matrix of this chain would consist of all zeros except for two entries which would be p and $1 - p$, the probability of a correct response and an error, respectively.

To incorporate the criterion into this approach we combine the

$$S = \sum_{j=K}^{M} \frac{M!}{j!(M-j)!}$$

sequences having K or more 1's into a single absorbing state, A. The remaining $2^M - S$ states are transient. Definition of these $2^M - S + 1$ states completes Step 1.

The initial probability of being in State A is simply the probability of reaching criterion on or before the Mth trial

$$\sum_{n=K}^{M} \frac{(n-1)!}{(K-1)!(n-K)!} p^K (1-p)^{n-K}.$$

The initial probability of the system starting in a transient state which contains i 1's and $M - i$ 0's is simply $p^i(1 - p)^{M-i}$.

The transition matrix of this chain consists of one row with the first entry equal to 1 and the remaining entries equal to 0, plus $2^M - S$ rows having all 0's except for two entries which are p and $1 - p$. For example, the criterion 2 out of 4 would have as its vector of initial probabilities and its transition matrix

$$\begin{bmatrix} a \\ p(1-p)^3 \\ p(1-p)^3 \\ p(1-p)^3 \\ p(1-p)^3 \\ (1-p)^4 \end{bmatrix} \quad \text{and} \quad \begin{array}{r|cccccc} & A & 1000 & 0100 & 0010 & 0001 & 0000 \\ \hline A & 1 & 0 & 0 & 0 & 0 & 0 \\ 1000 & 0 & 0 & 0 & 0 & p & 1-p \\ 0100 & p & 1-p & 0 & 0 & 0 & 0 \\ 0010 & p & 0 & 1-p & 0 & 0 & 0 \\ 0001 & p & 0 & 0 & 1-p & 0 & 0 \\ 0000 & 0 & 0 & 0 & 0 & p & 1-p \end{array}$$

where $a = 1 - 4p(1 - p)^3 - (1 - p)^4$.

Clearly the probability of reaching criterion on Trial $M + t$ is simply the probability that the system enters the absorbing state on Trial t. The initial probabilities for the transient states and two sets of recurrence relations given in Feller (1957, p. 363) determine the probability of absorption for each value of t.

We now develop, in the following paragraphs, the approximation method used to obtain the results for values of M greater than 10 (Tables 9–11). This method is used because the number of states in the transition matrix, $2^M - S + 1$, increases exponentially with M, giving transition matrices that rapidly become unmanageable even with a computer. The manageability problem is eliminated by using an approximate distribution function for absorption (i.e., attainment of criterion) as a function of trials. The derivation of this approximate distribution function rests upon the Approximation Assumption that the probability of being in a particular transient state, given that absorption has not taken place, remains equal over trials to the initial probability of being in that state.

Let $F(n)$ approximate the probability that absorption takes place on or before the nth trial of the process. It therefore approximates the probability of criterion being reached on or before Trial $M + n$ of the experiment. For absorption to occur on trial $n + 1$ of the process, three conditions must obtain:

1. Absorption has not occurred previously (the probability of this is $1 - F(n)$).

2. The system on Trial n of the process must be in a transient state having a 0 in the first (left hand) position which denotes an error on Trial $n - M + 1$ of the process, and exactly $K - 1$ 1's in the remaining $M - 1$ positions which denote $K - 1$ correct responses in Trials $n - M + 2$ to n of the process (the probability of this, given that condition i obtains, by the approximation assumption, is simply $(1 - p)b(K - 1, M - 1, p)$).

3. A correct response occurs on Trial $n + 1$ of the process (the probability of this is p).

We have, therefore, the approximate probability that the system is absorbed on Trial $n + 1$ of the process

$$F(n+1) - F(n) = D(1 - F(n)),$$

where $D = p(1 - p)b(K - 1, M - 1, p)$. The solution to this difference equation in $F(n)$ is

$$F(n) = 1 - (1 - F(0))(1 - D)^n,$$

where $F(0)$ is the previously obtained probability that the system starts in the absorbing state (i.e., that criterion was reached on or before Trial M of the experiment).

(Received August 20, 1964)

Psychological Bulletin
1965, Vol. 64, No. 1, 15–22

USE OF THE LIKELIHOOD FUNCTION IN INFERENCE[1]

D. A. SPROTT AND J. G. KALBFLEISCH

University of Waterloo

The likelihood function is defined and its use illustrated by a simple coin-tossing experiment. The distinction between the use of the likelihood function and the use of a test of significance is emphasized and illustrated by a simple genetics example. Some examples are given of experiments in psychology where the likelihood function is used to analyse the resulting data; the relative merits of the use of likelihood compared to other more standard methods of analysis are discussed.

The importance of the likelihood function has been emphasized for at least 40 years by Fisher (1922, 1925, 1934, 1946), and more recently specific examples of its use have been given (Barnard, Jenkins & Winston, 1962; Birnbaum, 1962; Fisher, 1955). In fact there is a "likelihood school" of statistics which says that all the information of the sample relative to the population under consideration is contained within the likelihood function and inferences should be based on it (Barnard et al., 1962; Birnbaum, 1962; Fisher, 1955; Sprott, 1961). The Bayesians (Edwards, Lindman, & Savage, 1963) also emphasize the importance of the likelihood function, but the use they make of it is different from that in this paper. Although the use and importance of the likelihood function are now being discussed in the statistical literature, there seems to be very little use or recognition of it in outside fields such as psychology. The purpose of this paper is to describe the likelihood function and its use in statistical inference. Section I defines and discusses the likelihood function and relative likelihood function and illustrates its use in inference by a simple example. Section II compares the likelihood approach with the somewhat similar, but logically distinct significance test. Section III gives an example from psychology of inferences based on the likelihood function from the data that would be difficult or impossible to assess accurately and rigorously in any other way. It is pointed out that perhaps the main drawback to the use of likelihood as a measure of uncertainty is that it is not probability and so is not intuitively obvious to some people. It has however a definite interpretation in terms of probabilities, and it should also be noted that very few methods of inference are capable of making probability statements about hypotheses or unknown parameters. The test of significance and confidence intervals certainly do not.

SECTION I: LIKELIHOOD

The use of likelihood is illustrated in the following example: a coin which turns up heads an unknown fraction, θ, of the time is tossed 10 times in order to obtain information about θ. Before the experiment is performed, the probability of observing x heads in 10 tosses for any specified value of θ is given by the binomial distribution

$$f(x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}$$

and can thus be calculated numerically for any x. After the experiment a particular value $x = x_0$ has been observed. We can now use $f(x_0|\theta)$ as a function of θ, employing the observed value, x_0, to rank possible values of θ according to their plausibilities. Suppose that $x_0 = 1$ head was observed. For any given θ the probability of the observed result is

$$f(1|\theta) = \binom{10}{1}\theta(1-\theta)^9. \qquad [1]$$

For example, if $\theta = .1$, the probability of observing one head is

$$f(1|.1) = \binom{10}{1}(.1)(.9)^9 = .387,$$

and if $\theta = .5$, the probability of observing

---

[1] This research was carried out by D. A. Sprott with a grant from the National Research Council of Canada.

The authors would like to thank R. K. Banks, R. V. Thysell, M. D. Vogel-Sprott, E. Tulving, R. G. Stanton, and J. C. Ogilivie for reading the manuscript and for comments, criticisms, and suggestions.

one head is

$$f(1|.5) = \binom{10}{1}(.5)(.5)^9 = .00977.$$

Thus under hypothesis H: $\theta = .1$ the result of the experiment is

$$\frac{.387}{.00977} = 40$$

times more probable than under H: $\theta = .5$. The data favor the hypothesis $\theta = .1$ 40 to 1 over the hypothesis $\theta = .5$. This is expressed by saying the likelihood of $\theta = .1$ is 40 times that of $\theta = .5$, or the likelihood of $\theta = .1$ compared to $\theta = .5$ is 40. Since hypotheses under which the observations are relatively probable will be considered more plausible and hence preferable to those under which the observations are relatively improbable we can use the function $f(1|\theta)$ to compare the plausibilities of different hypotheses in the light of the observation $x = 1$.

In general, before the experiment, $f(x|\theta)$, considered as a function of the observation x for a specified value of θ, yields the probability of observing x; after the experiment, $f(x|\theta)$, considered as a function of the unknown variable θ, and where x is set equal to its observed value, gives likelihoods of possible values of θ; $f(x|\theta)$ is called the likelihood function of θ and can be denoted by $L(\theta|x)$. Thus, before the experiment probabilities are relevant; after the experiment likelihoods are relevant (Barnard, 1949).

As can be seen from the above example, the likelihood is used to compare the plausibilities of different hypotheses under a given set of observations, so that only likelihood ratios are meaningful. In fact the likelihood ratio

$$\frac{L(\theta_1|x)}{L(\theta_2|x)}, \qquad [2]$$

which is the plausibility of $\theta = \theta_1$ versus $\theta = \theta_2$, is the ratio of the probability of the observations if $\theta = \theta_1$ to their probability if $\theta = \theta_2$. Since only ratios are used, the likelihood may be multiplied by any factor $C(x)$ independent of θ, for the ratio

$$\frac{C(x)L(\theta_1|x)}{C(x)L(\theta_2|x)} = \frac{L(\theta_1|x)}{L(\theta_2|x)}$$

remains unchanged. Thus in the above example the likelihood function for $x = 1$ can be taken to be

$$\theta(1-\theta)^9. \qquad [3]$$

Since it is desired to rank values of the unknown variable θ according to their plausibilities, special attention will center on the most plausible value or maximum likelihood estimate $\hat{\theta}$ of θ; the likelihoods of all other values of θ can be compared with that of the most plausible value θ by means of the likelihood ratio

$$R(\theta|x) = L(\theta|x)/L(\hat{\theta}|x). \qquad [4]$$

$R(\theta|x)$ varies between zero and one and gives the relative plausibility of θ compared to the most likely value $\theta = \hat{\theta}$ in the light of the observations. Any two values of θ can be compared by taking the ratio of their relative likelihoods, for

$$\frac{R(\theta_1|x)}{R(\theta_2|x)} = \frac{L(\theta_1|x)/L(\hat{\theta}|x)}{L(\theta_2|x)/L(\hat{\theta}|x)} = \frac{L(\theta_1|x)}{L(\theta_2|x)},$$

which by Expression 2 is the plausibility of $\theta = \theta_1$ versus $\theta = \theta_2$. In the first example, where $x = 1$ is observed, the maximum likelihood value $\hat{\theta}$ is $\frac{1}{10}$; inserting this into Expression 3 gives the maximum of the likelihood as

$$L\left(1|\hat{\theta} = \frac{1}{10}\right) = \frac{1}{10}\left(1 - \frac{1}{10}\right)^9 = \frac{1}{10}\left(\frac{9}{10}\right)^9, \qquad [5]$$

and the ratio $L(1|\theta)/L(1|\hat{\theta})$ of Expression 3 and Equation 5 is

$$R(\theta|1) = \frac{\theta(1-\theta)^9}{\left(\frac{1}{10}\right)\left(\frac{9}{10}\right)^9} = \frac{\theta(1-\theta)^9 10^{10}}{9^9}. \qquad [6]$$

By direct substitution into Equation 6, the values $R(\theta|1)$ can be calculated for various values of θ. Some results are given in Table 1. In Figure 1, $R(\theta|1)$ is plotted as a function of θ. All values of θ between .01 and .35 are fairly plausible. Values of θ smaller than .001 or larger than .5 are suspect, in that there are available hypotheses (e.g., $\theta = .1$) which are 40 times as likely. Values of θ as large as .7

are obviously extremely implausible in the light of the observation $x = 1$ and can be disregarded, as there are hypotheses under which our observation is 1000 times as probable. A discussion of a similar example will be found in Fisher (1955, pp. 66–73).

It can thus be seen that $R(\theta|x)$ is a measure of uncertainty or plausibility applicable to an unknown variable θ for given experimental observations x. Like probability, $R(\theta|x)$ varies between 0 and 1. Likelihoods are not probabilities however and the distinction is an important one to preserve. Likelihoods do not obey the rules of combination of probabilities. Although this is unimportant when only a single unknown variable, θ, is present, it becomes troublesome when there are many unknown variables, $\theta_1 \cdots \theta_k$.

TABLE 1

RELATIVE LIKELIHOODS FOR COIN-TOSSING EXAMPLE IN WHICH $x = 1$ WAS OBSERVED

θ	$R(\theta/1)$
.0	.0
.001	.0256
.01	.236
.05	.814
.10	1.000
.15	.897
.20	.683
.25	.486
.30	.313
.35	.187
.40	.105
.45	.0535
.50	.0255
.55	.0107
.50	.00406
.65	.00133
.70	.000346

SECTION II: LIKELIHOOD AND TEST OF SIGNIFICANCE

The standard or classical method of examining the foregoing example would be by means of the test of significance. To test the significance of $\theta = .5$, the probability of observing x less than or equal to 1 or greater than or equal to 9 is calculated:

$$\binom{10}{0}\left(\frac{1}{2}\right)^{10} + \binom{10}{1}\left(\frac{1}{2}\right)^{10} + \binom{10}{9}\left(\frac{1}{2}\right)^{10} + \binom{10}{10}\left(\frac{1}{2}\right)^{10} = \frac{11}{512} = .021.$$

Thus $H: \theta = .5$ is significant at the .021 level. It can be noted that the relative likelihood for $\theta = .5$ obtained in Section II for this example is .0255 (Table 1). The test of significance gives an alternative way of examining the evidence in this example. Several points should be noted however:

1. Neither the likelihood nor the test of significance is capable of making any probability statements about hypotheses or unknown variables. Under the latter, an hypothesis which is significant exactly at the 5% level does not necessarily have a 5% probability (or any other probability) of being true. Rather, an event as unfavorable as the one observed has a 5% probability of occurring if the hypothesis is true. The likelihood ratio yields statements of relative likelihoods of hypotheses; these are logically different from probabilities.

2. The test of significance makes no logical distinction between certain values not observed and the actual value observed. For instance, in the example the value $x = 0$, which was not observed, is utilized in the same way as $x = 1$, which was observed. Such tests have been criticized on this ground (Fisher, 1955). The likelihood, on the other hand, utilizes only the value of x actually observed, implying that other values of x not observed are irrelevant.

3. The likelihood function can be used

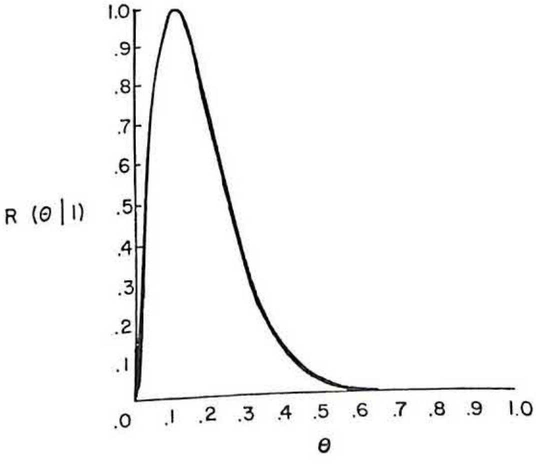


FIG. 1. Relative likelihood for coin-tossing example, $x = 1$ head observed.

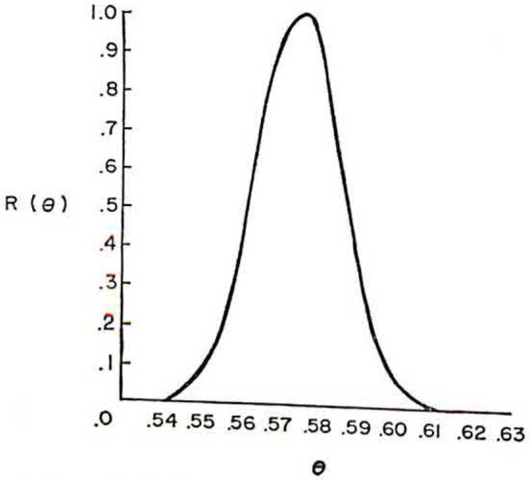


FIG. 2. Relative likelihood function for genetics example.

only to compare the plausibility of one hypothesis with that of another in the light of the data. In this sense the likelihood shows quantitatively the direction in which the weight of the evidence is pointing. The test of significance shows whether a specified hypothesis is consistent with the data. It should be noted that some hypotheses consistent with the data may still be relatively implausible. Indeed, the fact that an hypothesis is not significant merely means that no convincing evidence *against* the hypothesis has been demonstrated by that particular test; but there may be much more plausible hypotheses, which the relative likelihood will show. An example can show the relationship between a test of significance and the likelihood.

Example: If there are only two forms, M and N, of a gene at a given locus or place on the chromosome, then the genotypes MM, MN, NN are expected to occur with relative frequencies

$$\theta^2, \quad 2\theta(1-\theta), \quad (1-\theta)^2, \qquad [7]$$

respectively, if mating is at random with respect to these genes, where θ is the relative frequency of the M gene in the population[2] (cf. Li, 1955). Using the observations cited by Li (1955, p. 13) in which the frequency of the MM, MN, and NN blood types in a sample of 1,029 humans was observed to be 342, 500, and 187, respectively, the distinction between the use of a test of significance and the use of the likelihood can be illustrated. (*a*) Use of significance test—The significance test will indicate whether the Statistical Model 7 is consistent with the data at all. The probability of observing 342, 500, and 187 is given by the trinomial distribution with relative frequencies given by Model 7:

$$\begin{aligned}\Pr(342, 500, 187) &= \frac{1{,}029\,!}{342\,!\,500\,!\,187\,!}(\theta^2)^{342} \\ &\quad \times [2\theta(1-\theta)]^{500}[(1-\theta)^2]^{187} \\ &= \frac{1{,}029\,!\,2^{500}}{243\,!\,500\,!\,187\,!}\theta^{1184}(1-\theta)^{874}.\end{aligned}$$

Thus the likelihood of θ can be taken to be

$$\theta^{1184}(1-\theta)^{874} \qquad [8]$$

which attains its maximum at $\theta = \hat{\theta} = .5753$. Using this value for θ, the expected frequencies of MM, MN, NN in a sample of 1,029 are, by the theoretical frequencies in Model 7, $1{,}029(.5753)^2$, $1{,}029(2)(.5753)(.4247)$, $1{,}029(.4247)^2$ which are 340.6, 502.8, and 185.6. Comparing these with the observed frequencies 342, 500, and 187 by means of a chi-square test, and noting that 1*df* has been used to estimate the most likely value of θ (cf. Fisher, 1946), the resulting chi-square is .032 ($df = 1$). Since the probability of observing a chi-square larger than .032 is 85%, there is no evidence that the Model 7 is inconsistent with these data. (*b*) Use of the likelihood—Since the overall model above can be accepted as appropriate, the likelihood can now be used to rate the relative plausibility of values of θ. The maximum of the likelihood can be found by substituting $\theta = \hat{\theta} = .5753$ with Expression 8 to give

$$(.5753)^{1184}(.4247)^{874}.$$

The likelihood ratio obtained by dividing Expression 8 by its maximum in accordance with Equation 4 is

$$R(\theta) = \theta^{1184}(1-\theta)^{874}/(.5753)^{1184}(.4247)^{874},$$

and this measures the plausibility of θ relative to the most likely value $\theta = \hat{\theta} = .5753$. In Figure 2, $R(\theta)$ is plotted against θ. Values of θ outside the range (.54, .61) can be disregarded as the data overwhelmingly favor values inside the range.

---

[2] If the relative frequency of the M and N genes is θ and $1-\theta$, respectively, then the probability of having two M genes is θ^2, of having two N genes is $(1-\theta)^2$, and of having an M and an N gene in either order is $2\theta(1-\theta)$. This assumes that the acquiring of the M and N genes are independent events like the tossing of a coin.

Had the test of significance in *a* above yielded an improbably large value of χ^2, the entire Statistical Model 7 would have to be rejected as incompatible with the observations. Then the use of the likelihood as above would be irrelevant; it would be meaningless to ask about the relative plausibilities of values of θ in a model that was demonstrably incompatible with the data.

The test of significance can therefore be used to test the validity of the overall model, and, if the model is accepted as appropriate, the likelihood can be used to rate the relative plausibility of the unknown variables in it. The likelihood in fact indicates the direction in which the weight of the evidence is pointing once the hypothesis as a whole is found acceptable.

SECTION III: AN EXAMPLE FROM PSYCHOLOGY

The following data arise from an experiment by Banks and Vogel-Sprott (1964). The observations are the number of punished responses in 15 opportunities for response under four different delayed punishment times of 0 seconds, 30 seconds, 60 seconds, and 120 seconds. The results can be presented in a frequency table (Table 2) in which an entry n_{ij} in the body of the table gives the observed frequency of subjects in the *i*th delay-time group who gave *j* responses.

Specific questions under consideration are whether the number of responses depends on the delay-of-punishment time and whether delay times 30, 60, and 120 differ among themselves and from 0 delay. An analysis using number of responses as a continuous normal variate is questionable because the data are discrete and the distribution is U shaped and truncated at 15. A more appropriate procedure would be to use the above frequency table which contains counted (multinomial) data. The chi-square test of significance cannot be used because of small frequencies. The table could be condensed and classes combined, but much information would be lost. An extension of Fisher's (1946) exact test of significance for 2 × 2 tables could be used in principle, but even with a high speed computer this would not be feasible. Thus it is difficult to test whether the above hypotheses are actually inconsistent with the data (see Section II, 3). The use of the likelihood however is computationally simple and can throw some light on which way the experimental evidence is pointing.

TABLE 2

FREQUENCY OF PUNISHED RESPONSES

	Number of Responses							
Delay Time	1	2	3	4	5	6	7	8
0 seconds	2	3	0	2	0	2	0	0
30 seconds	0	1	1	1	0	0	0	1
60 seconds	0	1	0	1	0	1	1	0
120 seconds	0	1	0	0	0	0	0	0
Total	2	6	1	4	0	3	1	1
	9	10	11	12	13	14	15	Total
0 seconds	0	0	1	0	0	0	0	10
30 seconds	0	0	0	2	0	0	4	10
60 seconds	0	0	0	0	0	1	5	10
120 seconds	0	0	0	0	1	2	6	10
Total	0	0	1	2	1	3	15	40

Suppose p_{ij} is the hypothetical probability that a person makes *j* responses given delay time *i* ($i = 1$–4, $j = 1$–15). The hypothesis of independence between number of responses and delay time is mathematically

$$H_1:\quad p_{ij} = p_j \text{ independent of } i. \qquad [10]$$

The actual frequency of *j* responses under delay time *i* is n_{ij} (Table 2). The above data then follow multinomial distributions, and the likelihood function is the product of all $p_{ij}{}^{n_{ij}}$, which for notational convenience can be written as

$$\prod_{i,j} p_{ij}{}^{n_{ij}}. \qquad [11]$$

This example is somewhat more complex than the preceding examples, as the likelihoods of particular values of p_{ij} are not in question; rather we are concerned with the likelihoods of sets of p_{ij} satisfying Equations 10. In order to see if H_1 is reasonable, we can evaluate

TABLE 3

FREQUENCY OF PUNISHED RESPONSES

Delay Times	Number of Responses			
	1–5	6–10	11–15	Total
0 seconds	7	2	1	10
30 seconds	3	1	6	10
60 seconds	2	2	6	10
120 seconds	1	0	9	10
Total	13	5	22	40

the likelihood of the most likely set of values p_{ij} satisfying Equations 10, that is, maximize Expression 11 subject to Equations 10 and evaluate the relative likelihood of the resulting numbers $p_{ij} = \hat{p}_j$. If the most likely values of p_{ij} satisfying Equations 10 are unlikely, then all values satisfying Equations 10 are unlikely, so that H_1 itself is to that extent implausible.

The maximum likelihood values $p_{ij} = \hat{p}_j$ satisfying Equations 10 can be shown to be $\frac{n_{.j}}{n}$ where $n_{.j}$ is the total of Column j and n is the overall total. Substituting $p_{ij} = \hat{p}_j = \frac{n_{.j}}{n}$ into Expression 11 using the observations n_{ij} in Table 2, the resulting likelihood is

$$\left(\frac{2}{40}\right)^2\left(\frac{6}{40}\right)^6\left(\frac{1}{40}\right)\left(\frac{4}{40}\right)^4\left(\frac{3}{40}\right)^3\left(\frac{1}{40}\right)$$
$$\times\left(\frac{1}{40}\right)\left(\frac{1}{40}\right)\left(\frac{2}{40}\right)^2\left(\frac{1}{40}\right)\left(\frac{3}{40}\right)^3\left(\frac{15}{40}\right)^{15}.$$

The maximum of the likelihood is obtained for $\hat{p}_{ij} = \frac{n_{ij}}{n_{i.}}$ where $n_{i.}$ is the total of row i; this can be written, substituting in Expression 11

$$\prod_{i,j}\left(\frac{n_{ij}}{n_{i.}}\right)^{n_{ij}} = \prod_{i,j} n_{ij}{}^{n_{ij}} \Big/ \prod_{i} n_{i.}{}^{n_{i.}}. \quad [12]$$

The likelihood ratio or relative likelihood of $\hat{p}_{ij} = \frac{n_{.j}}{n}$ above is therefore the ratio of these two expressions, which, using the entries in Table 2 is

$$\frac{10^{40}2^2 6^6 4^4 3^3 2^2 3^3 15^{15}}{40^{40}2^2 3^3 2^2 2^2 2^2 4^4 5^5 2^2 6^6} = 4.9 \times 10^{-11},$$

ignoring 0's and 1's, since they contribute nothing to the products. The relative likelihood of the most likely set of values p_{ij} satisfying Equations 10 is extremely small so that H_1, even though it might not be contradicted by the data, is a relatively implausible hypothesis certainly not favored by the data. The evidence is pointing strongly away from H_1.

It may be thought that the delay times differ in respect of the rather erratic responses in the middle of the scale, that is, between 6 and 10. If responses 1–5, 6–10, 11–15 are combined, the result is Table 3. The hypothesis $H_1, p_{ij} = p_j$, can be tested in the same way, and the relative likelihood of the most likely values of p_{ij} under H_1 is

$$\frac{10^{40}13^{13}5^5 22^{22}}{40^{40}7^7 2^2 3^3 6^6 2^2 2^2 6^6 9^9} = 2.2 \times 10^{-4}.$$

It can be seen that even under this less sensitive classification the evidence does not favor the hypothesis that delay time produces no effect.

The hypothesis H_2 that delay times 30, 60, and 120 produce the same effect but possibly differ from 0 delay time is more complicated to test. Here the hypothesis is

$$H_2 \quad p_{ij} = p_j; \quad i = 2, 3, 4 \text{ only}.$$

The most likely values of p_{ij} satisfying H_2 can be shown to be

$$\hat{p}_{1j} = \frac{n_{1j}}{n}, \quad \hat{p}_{2j} = \hat{p}_{3j} = \hat{p}_{4j} = \frac{n_{.j} - n_{1j}}{n - n_{1.}} = \hat{p}_j.$$

The likelihood of these values of the p's can be obtained by substituting into Expression 11 using the numerical values in Table 3; the maximum of the likelihood is again found from Equation 12. The likelihood ratio arising from the results in Table 3 is therefore

$$\frac{10^{40}6^6 3^3 21^{21}7^7 2^2}{30^{30}10^{10}7^7 2^2 3^3 6^6 2^2 2^2 6^6 9^9} = .098.$$

Thus the evidence favors H_2, the equality of delay times 30, 60, and 120, over H_1, the equality of delay times 0, 30, 60, and 120, about 446–1 $\left(= \frac{.098}{2.2 \times 10^4}\right)$. Also the hypothesis H_2 is itself fairly plausible, so that it is reasonable to conclude that delay times 30, 60, and 120 produce essentially the same effect, the 0 time being responsible for the main differences observed.

The above tests exemplify the use of the likelihood in tables with entries too small to be attacked conveniently in any other way. Of course these tests, like the standard significance tests, are valid only if the hypotheses tested were considered relevant before the experiment was performed; they are not valid if the hypotheses tested were suggested by inspection of the data.

ANOTHER AREA OF APPLICATION

Section III dealt with an example in which the use of the likelihood appears to be the most convenient method of attack. A system in which there may be feedback of an unknown kind provides an example where the use of the likelihood function is the only method of attack. For instance, an observational variable y_t may be normally distributed about a mean βx_t with constant variance σ^2, where x_t is another observational variable at time t. If the x_t are independent of all y_t, standard regression analysis applies. If there is feedback of the form $x_t = y_{t-1}$, standard regression theory does not apply and the analysis using significance tests and probability (fiducial) intervals becomes complicated although still possible. If, however, there is feedback of an unknown form, $x_t = g(y_{t-1})$ where g is an unknown function, possibly independent of y_{t-1} (i.e., no feedback), then significance tests and probability intervals cannot even in principle be obtained as the required distributions will not be known. This kind of situation may possibly arise in experiments in which learning takes place. The likelihood function however will not depend on knowing the form of the function $g(y_{t-1})$; all that must be known is the numerical value of $x_t = g(y_{t-1})$. Whatever the form of $g(y_{t-1})$, the likelihood function based on a sample of n pairs of observations: x_1, y_1, x_2, y_2, $\cdots$, x_n, y_n is

$$\left(\frac{1}{\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum(y_t - \beta x_t)^2\right],$$

which attains its maximum at

$$\beta = \hat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2}, \quad \sigma^2 = \hat{\sigma}^2 = \frac{1}{n}\sum(y_t - \hat{\beta}x_t)^2.$$

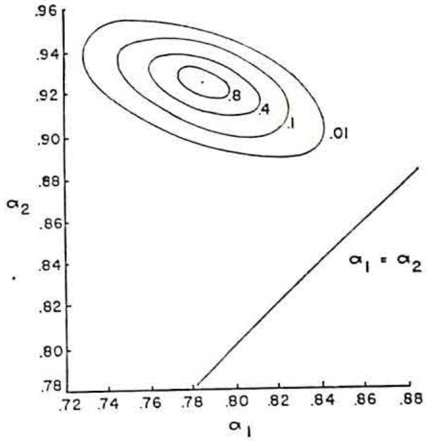


FIG. 3. Contours of constant likelihood for the Solomon-Wynne data.

Thus the relative likelihood can be calculated numerically for any pair (β, σ). For any value $\beta = \beta_0$, one can calculate the most likely value of σ^2 (i.e., σ^2 given β_0); if the relative likelihood of the resulting pair (β_0, σ^2) is small, then the value, β_0, itself is implausible, since no value of σ^2 will yield a pair (β_0, σ^2) with high likelihood. Examples of this kind are discussed in more detail by Barnard et al. (1962).

A final example of the use of the likelihood function is provided by the stochastic learning model of Bush and Mosteller (1955, pp. 238–249) applied to the Solomon-Wynne experiment. Using this model, the probability of a shock on Trial n when it is preceded by avoidances on k trials is

$$q_{n,k} = \alpha_1{}^k \alpha_2{}^{n-k},$$

where α_1 is called the avoidance parameter and α_2 the shock parameter. The likelihood function is therefore a product of terms, one for each trial, of the form

$$\alpha_1{}^k \alpha_2{}^{n-k} \quad \text{or} \quad 1 - \alpha_1{}^k \alpha_2{}^{n-k},$$

the former occurring in the product on shock trials, the latter occurring on avoidance trials. Since, in the Solomon-Wynne experiment, there are 30 dogs each with 24 trials, there are 720 factors in the likelihood function, making it too difficult to write down or compute on a desk calculator. The methods of Bush and Mosteller (1955), based on using suitable subsets of observations, could be used to write down an approximate likelihood function amenable to a desk calculator. However, it is relatively simple to program a high speed electronic computer to calculate exact likelihood values, using all of the data to obtain the maximum of the likelihood, and by Equation 4, the relative likelihood. This was done, and the results are shown graphically in Figure 3 in which contours of constant likelihood are plotted in terms of α_1, α_2, and the reference line $\alpha_1 = \alpha_2$ is shown. This graph represents a concise pictorial summary of everything the data have to say about α_1 and α_2. It can be seen from the elongated shape, for instance, that the data throw more light on α_2 than on α_1; that is, whatever be the value of α_2, plausible values of α_1 are in the range (.73, .840), whereas whatever be the value of α_1, the plausible values of α_2 are in the range (.89, .954)—the extremities of the 1% contour—the one range being about twice the size of the other. Also, the likelihood decreases very rapidly for small changes in α_1 and α_2, but about twice as rapidly for α_2. Thus, in terms of relative likelihood it is important for α_1 and α_2 both to be near the maximum. The region of appreciable likelihood is concentrated well above the line $\alpha_1 = \alpha_2$, so that there is little doubt that the evidence points strongly to α_2 being greater than α_1. The contours of constant likelihood also indicate how knowledge of one parameter will affect an inference concerning the other. For instance, if α_1 is known to be .746, the maximum of the relative likelihood is .1 and occurs at $\alpha_2 = .935$, so that values of α_2 in the range (.89, .954) now have a relative likelihood of at least $.01/.1 = .1$, rather than .01 as before. This approach can be compared with the more usual large sample approach used by Bush and Mosteller (1955) where the maximum likelihood estimates are obtained approximately as $\hat{\alpha}_1 = .797$, $\hat{\alpha}_2 = .923$ with asymptotic standard errors of about .04 and .015, respectively; they also give a theoretical interpretation of the results.

It can be seen from this and the preceding examples that the use of the likelihood has several advantages:

1. It represents a concise summary of all of the information in the data, and when graphed, it gives a pictorial summary of what the data have to say.

2. Inferences based on the likehood are exact for samples of all sizes; no assumption or mathematical investigation is required to ascertain if the sample is large enough to justify the method.

3. The rather complicated mathematics of asymptotic or large sample theory is not necessary, giving a relatively simple theoretical approach to inference.

Against these may be set the possible disadvantages of increased numerical computation in addition to the uncertainty being expressed (exactly) in terms of likelihood rather than (approximately in large samples) in terms of probability. Also, the likelihood method is not a test of significance and cannot be used to test the validity of a single hypothesis without reference to alternatives.

REFERENCES

Banks, R. K., & Vogel-Sprott, M. D. Effect of delayed punishment on an immediately rewarded response in humans. *Journal of Experimental Psychology*, 1965, **70**, in press.

Barnard, G. A. Statistical inference, *Journal of the Royal Statistical Society, B*, 1949, **11**, 116–149.

Barnard, G. A., Jenkins, G. M., & Winsten, C. B. Likelihood inference and time series (with discussions), *Journal of the Royal Statistical Society, A*, 1962, **125**, 321–372.

Birnbaum, A. On the foundations of statistical inference (with discussion), *Journal of the American Statistical Association*, 1962, **47**, 269–326.

Bush, R. R., & Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.

Edwards, W., Lindman, H., & Savage, L. J. Bayesian statistical inference for psychological research, *Psychological Review*, 1963, **70**, 193–242.

Fisher, R. A. On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society, A*, 1922, **222**, 309–368.

Fisher, R. A. Theory of statistical estimation, *Proceedings of the Cambridge Philosophical Society, Biological Sciences*, 1925, **22**, 700–725.

Fisher, R. A. Two new properties of mathematical likelihood, *Proceedings of the Royal Society, A*, 1934, **144**, 285–307.

Fisher, R. A. *Statistical methods for research workers*. London: Oliver & Boyd, 1946.

Fisher, R. A. *Statistical methods and scientific inference*. London: Oliver & Boyd, 1955.

Li, C. C. *Population genetics*. Chicago: Univer. Chicago Press, 1955.

Sprott, D. A. Similarities between likelihoods and associated distributions a posteriori, *Journal of the Royal Statistical Society, B*, 1961, **23**, 460–468.

(Received September 16, 1964)

Psychological Bulletin
1965, Vol. 64, No. 1, 23–51

PROPERTIES OF ORGANIZATION STRUCTURE IN RELATION TO JOB ATTITUDES AND JOB BEHAVIOR[1]

LYMAN W. PORTER AND EDWARD E. LAWLER, III
University of California, Berkeley *Yale University*

This article reviews the results of empirical field studies that have investigated the relationships between properties of organization structure and job attitudes and job behavior in business and industrial organizations. The following 7 structural properties were examined: organizational levels, line/staff hierarchies, span of control, subunit size, total-organization size, tall/flat shape, and centralized/decentralized shape. At least 5 of these 7 variables (with the possible exceptions being span of control and centralized/decentralized shape) were found to be significantly related to one or more attitude or behavioral variables. Implications of these findings for organization theory and future research are discussed.

All organizations are structured, in the sense of having positions and parts which are systematically related to other positions and parts. Since organizations vary in their structure, it is appropriate to examine the question of whether differences in the structure of organizations are related to differences in the attitudes and behavior of their members. The purpose of this article is to review the results of published research (prior to 1964) relevant to such relationships existing in business and industrial organizations.

Although organizations have been the focus of research and interest in sociology since the time of Weber's first writings on bureaucracy in the 1920s, and in industrial psychology since the time of the Hawthorne studies in the early 1930s, it has been only in the past decade or so that intensive and concentrated attention has been given to organizations in the behavioral sciences. The major development in this area in psychology over the past 10 years has been the advent of so-called "modern organization theory." Such theory, as developed by Likert, Haire, McGregor, Argyris, and others, has been advocated as a contrast with, and improvement on, so-called "classical organization theory" as developed by Fayol, Urwick, Taylor, and others. Whereas the classical theorists, with their discussions of chain of command, specialization of function, span of control and so forth, placed heavy emphasis on factors related to organization structure, the modern psychological theorists, such as those mentioned previously, have paid relatively little attention to the structural aspects of organizations. Bennis (1959) aptly summed up this point several years ago when he pointed out that classical theorists talked about "organizations without people," while modern theorists often seem to talk about "people without organizations." Instead of emphasizing the structure of organizations, latter-day theorists have advocated supervisory human relations training, *T* groups, attention to the nonfinancial needs of the employees and the like, as methods for improving organizational administration and leadership. The efficacy of such methods, however, may well be reduced by failure to take into account the structural characteristics of organizations.

Not only have organization theorists tended to ignore structural variables, but also so have many of the researchers in this area. Despite the apparent lack of concentrated empirical efforts to understand the effects of structural variables (other than the variable of work-group size), there are, nevertheless, a number of studies in the psychological, sociological, and management-science literature that are

---

[1] The authors are indebted to the following individuals for their comments and suggestions regarding earlier versions of this paper: Chris Argyris, Peter M. Blau, Edwin E. Ghiselli, Thomas M. Lodahl, George Strauss, and Victor H. Vroom. We should also like to acknowledge the financial support of the Institute of Industrial Relations, University of California, Berkeley.

pertinent to a consideration of structure. Although many of these studies were not conceived by their authors as investigations pertaining to structure as such, they do provide relevant and important evidence and thus will be included in this review.

For the purposes of this paper, structure has been defined to mean the positions and parts of organizations and their systematic and relatively enduring relationships to each other. Thus, we are referring to the formal structure of organizations as might be indicated, in a superficial way at least, in the formal "organization chart." Within the boundaries of this broad definition of structure, we have identified the following seven structural properties or variables:

Suborganization properties
1. Organizational levels
2. Line and staff hierarchies
3. Span of control
4. Size: subunits

Total-organization properties
5. Size: total organizations
6. Shape: tall or flat
7. Shape: centralized or decentralized

The first four variables above can be considered as suborganization properties of structure inasmuch as they permit comparisons of positions or parts contained within organizations with other positions or parts, either contained within the same organizations or within other organizations. The last three variables can be considered total-organization properties of structure because they require comparisons among organizations as complete entities. This sort of distinction between suborganization and total-organization types of properties will be shown to be especially crucial in the discussion of the effects of "size." It is also important to stress that these seven properties or dimensions of structure should not be considered mutually exclusive. For example, "span of control" has an obvious connection to the variable of "tall/flat" shape of total organizations. Although such variables are not clearly distinct from each other, we have discussed them separately since they are typically considered separately in both the research and textbook literature on organizations.

In essence, for this review these seven properties of organization structure will be regarded as the "independent" variables. Since almost none of the studies reviewed were experimental studies in the strict use of the term "experiment," any relationships found between structural variables and job attitudes or job behavior cannot be considered "cause-effect" relationships. Thus, neither the studies nor our conclusions from them can prove that the variations in structure cause the particular attitudes or behavior. However, the structural properties are referred to as the independent variables in the sense that if there are any cause-effect relationships present they probably are due either to the variation in structure causing the variation in attitudes or behavior, or else to some third variable determining both the structure and the attitudes and behavior. Although it is conceivable that the kinds of attitudes or behavior usually investigated in these studies could determine the structure of organizations, this seems unlikely (at least in a direct fashion) in most cases.

Turning now to our "dependent" variables, job attitudes and job behavior, we have used the term "attitude" in its broadest and loosest sense of "opinion concerning some object," in this case jobs or aspects of jobs. Included under job-attitude studies would be the typical morale study, where the specific attitude questions are concerned with opinions about the organization, the supervisors, the working conditions and other factors directly connected with the job situation in which the employee operates. Excluded from the review are studies focusing on attitudes toward the self, such as self-description and vocational-interest studies, on the assumption that they are more likely to describe what groups or types of people are like rather than what positions or parts of organizations are like.

Considered under the heading of job behavior are studies that were (with a few exceptions) concerned with the following types of information: performance and output rates or ratings, turnover rates, absenteeism rates, accident rates, and employee-grievance rates. In this review the problem of the relationship between job attitudes and these various indices of behavior have not been dealt with;

these have merely been considered to be two categories of dependent variables that are of interest to social scientists studying organizations.

Finally, the three major criteria for the inclusion of studies reviewed can be listed:

1. Studies of industrial and business organizations. The review is purposely restricted to studies dealing with these types of organizations—rather than also including other kinds of organizations such as governmental agencies, military units, schools, churches, and the like—in order to keep the discussion reasonably focused and delineated.

2. Comparative studies. Only those studies which have dealt with at least two (preferably, three or more) positions or degrees of structure along a given property or dimension are included. Thus, for example, a study of only foremen's attitudes would not be included in the section on organizational levels since such a study did not compare foremen's attitudes with those of employees at some other organizational level. Likewise, a study of only small companies would not be included in the section on studies of total size, since no comparisons within that study are possible between small organizations and other-sized organizations. Under this criterion, case studies of single organizations were also excluded.

3. Empirical studies. With very few exceptions, only those studies providing empirical data are included. Articles based wholly or mostly on the opinions of the author unsupported by actual data have been excluded. In addition, articles based on casual observations where it seemed unlikely that the procedures could be easily repeated by other researchers have also been excluded.

The sections of the review to follow are arranged in the same order as the previously-listed set of seven properties of organization structure. Thus, we first take up properties permitting suborganization comparisons, followed by properties concerned with total-organization comparisons.

ORGANIZATIONAL LEVELS

For many years social scientists interested in industrial organizations have tended either to ignore the effect of organizational levels or else to concentrate on the simple dichotomy between managers and workers. In the 1940s most textbook writers in the area of human relations, following the lead of the Hawthorne investigations, spotlighted the plight of workers vis-à-vis managers in terms of the types of physical, social, and psychological conditions they had to endure. The emphasis was almost completely on the worker-management comparison. This tradition carried over into the 1950s when the "modern" management theorists, such as Argyris (1957), Haire (1956), Leavitt (1958), Likert (1961), and McGregor (1960), continued to focus, for the most part, on the relationship of managers to nonmanagement subordinates.

This relative lack of attention to the effects of differentiation among levels within management is somewhat surprising when one considers the amount of interest that industrial psychologists and sociologists have shown in the effects of division of labor. Almost always, however, division of labor has been studied as it occurs along a horizontal dimension at the rank-and-file worker level. The fact that there is not only a division of labor horizontally but also vertically within industrial organizations has frequently been neglected. The importance of paying closer attention to this vertical dimension of structure has been well stated by Pfiffner and Sherwood (1960) when they point out that:

> [The] differentiation of task between echelons is of more significance to the selection and training of leaders at the several levels than may be indicated by the attention accorded it in the past. The psychological adjustment necessary when one goes from one level to another is often difficult because of the tendency to continue former behavior patterns . . . . At first glance this might appear to be a problem of human relations rather than of formal organization, but such a conclusion would be only partially true. Good job descriptions should reflect task differentiation at the various echelons. It is a matter of tasks combined with behavior . . . [p. 139].

In surveying the literature pertinent to organizational levels and their impact on, or relationship to, job attitudes and behavior, we encounter several difficult boundary or definitional problems. The first such problem concerns the question of whether studies of *occupational* levels should be included in the review. For our purposes, since the focus is on *organizational* structural variables, such occupational level studies will be generally omitted on the basis that they ordinarily investigate levels in a societal or cultural setting rather than levels within an organizational setting. A second problem in defining studies of organizational levels is posed by the investigation directed toward superior-subordinate relationships. Although in one sense these are studies that compare two different organizational levels, they in fact seldom provide data that contribute to our knowledge of the effects of the vertical dimension of structure. This is because such studies are usually concerned with implications for superior-subordinate authority relationships and not with implications for attitudes or behavior that may be differentially required or exhibited at various *job levels* within organizations. Most such studies, therefore, will not be covered in this review.

Within these broad limits, this section of the review will analyze the results of studies containing two or more organizational levels as the independent type of variable and one or more categories of directly job-related attitudes or behavior as the dependent type of variable.

Attitudes

Job Satisfaction and Morale. The first studies to be considered will be those that compare management with nonmanagement employees; after that, studies dealing with comparisons among levels within management will be reviewed.

Herzberg, Mausner, Peterson, and Capwell (1957) have summarized the literature through 1954 relevant to job satisfaction attitudes associated with organizational levels. They conclude that "one unequivocal fact emerges from the studies of job satisfaction; the higher the level of occupation, the higher the morale [p. 20]." It is clear from this statement that their review included studies of occupational level. The problem of drawing conclusions from these studies is that frequently the occupational level categories include people who do not work in industrial or business organizations (e.g., professional people included in a category of "executive and professional"). Nevertheless, the Herzberg review does cite five studies "in which the morale of workers was compared to that of their supervisors (Ash, 1954; Browne & Neitzel, 1952; Campbell, 1948; Hull & Kolstad, 1942; Kolstad, 1944)." The study by Browne and Neitzel was actually based on different supervisory levels and will thus be reviewed below. The other four studies, however, did show greater job satisfaction among supervisory personnel when compared to workers.

In a study not included in the Herzberg review, Morse (1953) reported on the satisfactions of 61 supervisory personnel compared with 600 workers in a single company. She found that:

> The supervisors are considerably more satisfied with their jobs and with the company as a place to work. They are somewhat less satisfied than the employees with their salaries and are about equal in satisfaction with the employees regarding the advancement they have received in the company [p. 98].

Thus, Morse's conclusions are in general agreement with the other studies cited above, although, as she points out, first-level supervisors may not be as satisfied as the rank and file with the specific aspect of wages.

Since the publication of the Herzberg review, only one study (Handyside, 1961) has appeared that has focused on management versus nonmanagement differences in job satisfaction. This study involved 30 managers and supervisory personnel and 467 production workers and found that "median satisfaction" (toward jobs) was higher for managerial personnel.

In recent years the content of job satisfaction studies of different organizational levels seems to have shifted from an exclusive focus on comparisons of supervisory personnel versus rank-and-file workers to comparisons among different levels of management within organizations.

A major point at issue in these studies of levels within management seems to be whether satisfaction increases with each higher level of management, or whether middle levels may have poorer morale than either lower or upper managerial levels. In their review of the literature through 1954, Herzberg et al. (1957) state that: "In large concerns, middle levels of management often have very poor morale [p. 23]." Their review refers to two articles relevant to this point (Benge, 1944; *Fortune*, 1950), but neither article presents enough data on either the samples of respondents or the results to draw any meaningful conclusions. Thus, there seems to be no substantial basis in the literature prior to 1954 to support this conclusion.

Recent studies, plus one appearing prior to the Herzberg review, seem to be nearly unanimous in concluding that job satisfaction or morale does increase monotonically with increasing levels of management, and that therefore middle managers are more satisfied than those below them in the organization but less satisfied than those above. The first such study, by Browne and Neitzel (1952), found that morale scores for three levels of supervision were "positively related to the echelon level of the supervisors [p. 90]." In 1961 two articles reporting on different levels of management appeared (Porter, 1961; Rosen, 1961a). Rosen's study, which was concerned with satisfaction with conditions of work, presented findings for three levels of management within a single plant. He found that the top managers in the plant and the middle managers did not differ significantly from each other in satisfaction, but both groups of managers were significantly more satisfied than first-level supervisors. Rosen also noted, however, that there was a similarity of profiles of satisfaction among the three levels of management, such that there were high rank-order correlations (ranging from .67 to .93) between each pair of levels in terms of the degree of satisfaction with the 24 items.

In Porter's study (1961), two levels of management, first-line supervisors and lower-middle management, were surveyed in three different types of companies with a questionnaire designed to tap satisfaction of different types of needs relevant to a Maslow hierarchical classification of needs. The results showed that the higher of the two management levels generally reported greater satisfaction and, as was the case in Rosen's investigation, the pattern of need satisfactions tended to be similar between the two levels.

Subsequent to these 1961 studies by Porter and Rosen, Porter has completed a large-scale investigation of managerial need satisfactions across all levels of management, from first-level supervisors to company presidents. This study (Porter, 1962, 1964) sampled more than a thousand managers from a wide variety of different types and sizes of companies located throughout the country. The results were in general agreement with those from the previous Porter and Rosen studies. Satisfaction increased with each higher level of management for three of five needs (esteem, autonomy, and self-actualization needs), and profiles of satisfaction were generally similar across levels. (For security and social needs, satisfaction was roughly equal across all levels of management.) It should be emphasized that satisfaction in this study was measured by the difference between obtained and expected fulfillment. This is important because it means that for at least three types of needs lower-level managers were not just getting less fulfillment, as might be obvious, but there was a larger difference between their expectations and their fulfillment than was found at upper levels of management.

Much of the same sort of results were reported in a study published by the Opinion Research Corporation (1962). Attitude data were collected from 1,200 managers showing that the percentage of unfulfilled needs decreased from lower to middle to top management. For almost all of the specific questions asked of the respondents there were no reversals wherein a higher level of management expressed greater dissatisfaction than a lower level.

Finally, in a study as yet unpublished in detail, Haire, Ghiselli, and Porter (1963) found that in a cross-cultural investigation of managerial attitudes in 14 countries, higher levels of management on the whole reported greater degrees of need satisfaction in their jobs than did lower-level managers. It thus appears that the phenomenon of increased job satisfaction with increasing level of management is not confined to the situation existing in American companies but tends to be a worldwide fact of industrial organizations.

To summarize briefly at this point, it can be stated with some degree of assurance that the available literature on job satisfaction across different levels of organizations shows increasing job satisfaction at each higher level. This is true both for nonmanagement versus management comparisons and also for comparisons within the management parts of hierarchies. In addition, one other finding appears with some consistency: Patterns of satisfaction are roughly similar across different organizational levels, at least within management.

Other Attitudes. A number of studies have been concerned with the relationship of organizational levels to attitudes other than degree of satisfaction. However, such studies are somewhat difficult to classify and summarize since they cover a great diversity of types of attitudes.

There are, nevertheless, several groupings that are reasonably discernible. The first of these comprises studies of the relative importance of different types of needs and motives at different organizational levels. Two studies (*Fortune,* 1947; Kahn, 1958) have made direct comparisons between nonmanagement and management levels with regard to the importance of different motivating factors. The two studies seem to present somewhat conflicting results for these comparisons. Kahn's findings indicated that both workers and their supervisors attached high importance to security, but that supervisors attached less importance to high wages than did the workers. The *Fortune* study, on the other hand, indicated that when security was pitted against high wages in an "either-or" choice situation, high-level executives chose the high wage situation. It is obviously difficult to compare the two sets of findings since both the samples and the procedures differed. What is clear is that we need considerably more evidence from a wide variety of types of organizations before we can make any broad generalizations concerning manager-worker differences in the importance attached to different motivating factors.

Several studies in the past 5 years have made comparisons of the importance attached to different needs and conditions of work among different levels within management. The first of these studies was by Rosen and Weaver (1960) who studied the degree of importance attached to 24 conditions of work by three levels of management within one plant. They found high rank-order correlations (ranging from .76 to .90) between each pair of levels with regard to the degree of importance attached to each of the 24 items. In terms of overall importance, the first-level supervisors attached generally more importance to most of the items. Subsequent to the Rosen and Weaver study, Porter published two articles dealing with the importance attached to different psychological needs by different managerial levels (Porter, 1961; 1963a). Both studies confirm one aspect of the findings of Rosen and Weaver; namely, different managerial levels tend to be similar with regard to what they consider most important. However, the later and more comprehensive of the Porter studies differed in one respect from both the earlier Porter study and the Rosen and Weaver study. It showed that higher-level managers tended to attach somewhat more rather than less importance to various needs, whereas the earlier studies indicated that higher managerial levels regarded most needs or conditions as less important compared to lower levels. The difference in these findings may lie in the fact that in the later Porter study the degree of formal education of the respondents was about equivalent across management levels, while in the earlier studies the degree of education was either not controlled or else the lower-level respondents had less education. It is conceivable that employees with less education may be more prone to consider everything as important, whereas those with greater formal schooling may be somewhat more discriminating in attaching importance to different needs and conditions. Therefore, it may be necessary to control for differences in formal education when comparing different management levels with regard to the importance attached to different psychological needs or conditions of work.

Finally, the cross-cultural study by Haire, Ghiselli, and Porter (1963) showed that for almost all of the countries in their sample there was again a strong similarity between upper-level and lower-level managers in the relative importance attached to different needs.

Turning to a different area of job attitudes, we find a small group of studies that can be put under the general heading of role perception studies (Coates & Pellegrin, 1957; Porter & Henry, 1964a; Rosen, 1961b; Triandis, 1959b). Although each of these studies was in some way concerned with role perceptions, it should be noted that they were addressed to rather different aspects of role perceptions, they utilized different types of samples, and, particularly, they employed quite different methods to obtain their data. For instance, Triandis (1959b) utilized the semantic differential technique to study the "differential perception of certain jobs and people" by upper-level managers, lower-level managers, and clerks and workers from one company. He concluded that "the most significant variable in the perception of jobs is the *level* of the job [being perceived; p. 224]." In addition, he also found that managers made finer discriminations among their job perceptions than did workers. Rosen (1961b) investigated how three levels of management in a particular company ranked 16 "role prescriptions." His findings showed that managers at the two lower levels tended to be similar in how they evaluated their role as supervisors, but each level differed somewhat from the highest of the three levels. Porter and Henry (1964a) studied the perceptions of managers in terms of the type of behavior each level believed was most important for jobs at their particular level, and found that at each higher level of management increasing emphasis was put on "inner-directed" (as opposed to "other-directed") personality traits. Coates and Pellegrin (1957), in a study directed primarily toward self-perceptions, found that both supervisors and subordinates tended to "be aware of the rewards, penalties, and sacrifices associated with high-level roles. Such an awareness differentially influences achievement desires on the two levels [p. 220]."

An overall look at these diverse studies of role perception among different organizational levels reveals that again the variable of level seems to have a strong relationship to these types of perceptions, just as it did to perceptions of need and job satisfaction.

Two investigations have compared the factorial structures of attitudes of lower-level personnel versus those at higher levels. Baehr (1954), factoring the Science Research Associates (SRA) Employee Inventory for a "high-status" group of managers and white collar personnel and a "low-status" group of factory workers and routine clerical personnel, found considerable, but not complete, similarity in the factors emerging from the two groups. What differences there were indicated that the factors of the high-status group tended to be more concerned with the overall functioning of the organization while those of the low-status group were more relevant to specific aspects of the immediate work situation. Triandis (1960) factored the semantic job structures of managers and supervisors in a manufacturing plant and compared these with the factored job structures of workers in the same plant. He found even greater factorial similarity between the two groups than did Baehr and concluded that: "Management and workers, one might expect, will differ in their factorial structures when the domain of jobs is considered. The answer is that this is not true. The factor structures are surprisingly similar [p. 300]." It appears, therefore, from these two studies, that organizational level does not have a strong relationship to the structure of job attitudes.

The salience of different features of jobs and jobholders appears to be related to the level at which a person is working in an organization. Two studies which demonstrate this are another article by Triandis and one by Pellegrin and Coates. Triandis (1959a) found that upper-level managers focused on such "expressive" categories of description as "graciousness" and "polish," while lower-level managers and workers emphasized "instrumental" categories such as "intelligence," "skill," and "dependability," in describing different groups of employees. Likewise, he also found differences associated with the three levels of respondents when they were describing different jobs. The two managerial groups tended to use the criterion of the "nature of the work" to distinguish among different jobs, whereas the workers made their distinctions among jobs primarily on the basis of job requirements, particularly intelligence. Pellegrin and Coates (1957), in interviews with a sample of 50 "top-level" executives and 50 "first-level" supervisors, found that the former emphasized the intrinsic features of their jobs whereas the latter viewed their jobs more as means to off-job rewards. This may indicate, in the terms of Herzberg, Mausner, and Synderman (1959), that managers in

TABLE 1

STUDIES OF RELATIONSHIPS OF ORGANIZATIONAL LEVELS TO JOB BEHAVIOR

Investigators	Type of behavior	*N* of individuals	*N* of levels	*N* of companies
Burns (1954)	Interactions and activities	4	3	1
Davis (1953)	Communications (grapevine)	65	6	1
Dubin & Spray (1964)	Interactions and activities	8	2	3
Martin (1959)	Decision making	?	4	1
Shartle (1956)	Activities	14	2	7

higher-level positions focus more on "motivating" aspects of their positions while supervisors in lower-level jobs are oriented more toward concern with the "hygienic" conditions of their positions.

One other area of job attitudes in relation to organizational levels has received some attention, namely, types of leadership attitudes at different levels. The Ohio State studies (e.g., Fleishman, 1953; Fleishman, Harris, & Burtt, 1955) have demonstrated that attitudes about how supervisors should behave in relation to their workers vary depending on the organizational level of the respondent. For example, Fleishman's (1953) article showed that the higher the level of a manager, the less likely he was to believe that a supervisor should show "Consideration" attitudes toward his subordinates and the more likely he was to think a supervisor should put emphasis on "Initiating Structure" in dealing with subordinates.

Behavior

Studies of the relationships between organizational levels and individuals' behavior are relatively infrequent. The nature of five such studies that were located in the literature, including the sizes of their samples, is indicated in Table 1. This table shows that the sampling, in terms of both individual subjects and organizations, is extremely limited. Any conclusions emerging from these studies, therefore, must be regarded with great caution. One other point should be noted in Table 1. These five studies dealing with behavior in relation to organizational level are not concerned with rates of production, absenteeism, safety, turnover, and the like. Instead, they concentrate on interpersonal interaction patterns, types of decisions, and types of functional activities performed by the individual in his job. The reason these studies do not focus on the types of behavior usually studied in the typical field investigation is that they are all concerned with differences in behavior between or among levels of management. Thus, output or production in the usual sense is not easily measured. Also, it is probably assumed that turnover, absenteeism, and accident rates are so negligible at managerial levels that there would be little or no variation in these indices from level to level.

Keeping in mind the severely restricted nature of the samples in these five studies, there are several conclusions that seem to be supported by the data so far obtained:

1. Informal Communication (Informational) Patterns. Davis' (1953) study of the grapevine operating in the management sector of a particular company indicated that higher-level managers are much better informed on informal communication (as well as, presumably, on formal communication) items than lower-level managers and supervisors. Davis stressed the fact that, in this company at least, first-level supervisors were the most isolated from news traveling through the informal communication channels.

2. Interpersonal Interaction Patterns. Three major types of interaction variables have been examined in several studies: (*a*) The ratio of upward (or superior) contacts to downward (or subordinate) contacts—the three studies (Burns, 1954; Dubin & Spray, 1964; Martin, 1959) that deal with this variable present contradictory findings. Burns' data (as pointed out by Dubin, 1962) show that the ratio of superior to subordinate contacts decreases the higher the level of management; Dubin and

Spray's data seem to indicate the opposite; Martin's results show no changes in the ratio in relation to level. (*b*) Percentage of peer contacts—both Burns' and Martin's findings indicate that the relative number of contacts with peers decreases the higher the level of management. Dubin and Spray's data are inconclusive on this point. (*c*) Percentage of contacts with individuals outside the organizational unit (e.g., factory, office, or company)—the data of both the Martin and the Dubin and Spray studies show a definite trend for such outside contacts to increase the higher the level of management, as one might expect. Burns' study indirectly supports this point, as does one by Shartle (1956), who found that the heads of seven organizations spent a greater proportion of their time on public relations activities than did their immediate assistants. (*d*) Percentage of direct (face-to-face) contacts—Martin found the proportion of face-to-face contacts decreasing at higher management levels.

3. Types of Decisions. Martin's study is the only one of the five studies to focus on the differences in the types of decisions made at different levels of management. In his investigation of four levels (from works manager down to shift foreman) within a single plant, he found that the higher the managerial level, (a) the greater the "duration" of the decision (i.e., increase in time between the inception of the decision process and the actual implementation of this decision and the verification of its correctness or incorrectness), and hence the greater the increase in the "riskiness" of the decision, and (b) the more abstract the decision (i.e., an increasing lack of "structure" of the data relevant to the decision).

4. Types of Activities (e.g., planning, coordinating, production, etc.). Data concerning differences in patterns of activities at different levels of management are so meager that generalizations cannot be made in this area, even though several studies (Burns, 1954; Dubin & Spray, 1964; Shartle, 1956) have attempted to collect such information.

Summary

A review of the relevant literature demonstrates that organizational levels are strongly related to both attitudes and behavior. The evidence shows that perceived job and need satisfactions increase not only from rank-and-file positions to managerial positions, but also from lower management positions to middle- and upper-level positions. Thus, level seems clearly to be related to what individuals think they are receiving from their jobs. Level also seems to show a relationship to what employees think they must put into their jobs, in terms of the behavior they perceive to be required by their jobs. On the other hand, it should also be stressed that level does not necessarily affect all types of job attitudes. For example, it seemed to have only a slight relationship to the relative importance attached to different kinds of needs.

In terms of behavior that is actually exhibited by employees, especially within managerial ranks, level seems to affect the amount of information a person receives in his job, the types of interpersonal relationships he has on his job, and the types and nature of the decisions he must make in his position. However, the evidence concerning these behavioral variables in relation to level is limited, and hence conclusions concerning them must be regarded as tentative at the present time.

The major research trend in this area in recent years, both in studies of attitudes and of behavior, seems to be a shift toward an increasing focus on the differentiation of organizational levels within management in contrast with the former (almost complete) concentration on management versus nonmanagement differences. Eventually, it may be possible to describe organizational differences in major aspects of attitudes and behavior for at least four discrete points along the hierarchical or level dimension of structure: rank-and-file positions, first-line supervision, middle management, and top management. To date, however, we have no major sets of attitudinal or behavioral data from a specific study that adequately sample these four levels across different kinds of industrial or business organizations.

Future research on organizational levels will need to focus on the question of what it is about different levels that seems to create different job attitudes and job behavior. In other words, "organizational level" is merely a convenient label for other, more fundamental, variables that must be operating in the situation. An understanding of the nature of these other variables and how they are distributed along the vertical dimension of structure is a research task for the future in this area.

Future research will also need to focus on the necessity of disentangling the effects of individual-difference variables from the effects of organizational-level variables. Practically all of the studies so far carried out in this area have failed to consider the fact that individuals at different organizational levels may vary systematically in intelligence and/or personality traits. To what extent the obtained differences in attitudes and behavior are a function of these types of differences relative to differences due to job level is an open question. It is a question that will require sophisticated research for an answer.

LINE AND STAFF HIERARCHIES

The distinction between line and staff hierarchies is an old one in the study of formal organizations. Traditionally, positions concerned with the main operations of the organization and within the direct chain of command have been considered part of the line hierarchy, while positions concerned with auxiliary services and outside the direct chain of command have been considered part of the staff hierarchy. The staff's function is to provide specialized aid to the line, which is composed of "generalists." Recently, this traditional distinction has received attention from two sources. One group of writers (Brown, 1954; Fisch, 1961; Leavitt & Whisler, 1958; McGregor, 1960) has contended that the distinction between these two parallel hierarchies is disappearing under the impact of modern technology. A second group of writers has dealt with the type of relationships that exist between line and staff managers. Both Dalton (1959) and Whyte (1961) have illustrated the type of conflicts that can develop between managers in these two hierarchies existing within organizations.

Attitudes

The evidence available concerning the amount of satisfaction provided by line and staff positions is consistent in showing line managers to be more satisfied than staff managers. Three studies by the Opinion Research Corporation (1959) reported that engineers and scientists in staff positions were less satisfied with their jobs than engineers and scientists in line positions. In more recent studies using broader samples of managers, Rosen (1961a) and Porter (1963b) found that line managers reported a greater incidence of desirable conditions of work and perceived greater need satisfactions, especially in esteem and self-actualization need areas.

Three studies compared line and staff managers on the amount of importance they attached to job-related factors and needs, and all three found essentially no differences between the two groups of managers. Rosen and Weaver (1960) found that managers in line and staff positions did not differ substantially in their ratings of the importance of 24 desirable conditions of work. Porter (1963b) found that there were no line-staff differences in the importance attached to four of five need areas; only on the need for autonomy did the two groups differ, with staff managers feeling this need was more important. Lawler and Porter (1963) found no differences between the two types of managers in the importance attached to pay.

In a study involving a somewhat different type of attitude, Porter and Henry (1964b) asked line and staff managers to rank five other-directed and five inner-directed personality traits on the basis of how important they are for success in their management positions. The results showed that staff managers, as Whyte (1956) and others might have predicted, felt they have to show more other-directed behavior to succeed in their jobs than line managers.

Behavior

Only three studies were located that permitted comparisons of line and staff personnel on behavioral indices. A frequently cited study by Dalton (1950) found that staff managers in three plants had a turnover rate between two and four times that of line managers. Davis (1953), in a study of one company, reported that staff managers were better informed than line managers. He attributed this superior knowledge on the part of staff managers to their greater mobility.

> Staff executives in such areas as personnel and control found that their duties both required and allowed them to get out of their offices, made it easy for them to walk through other departments without someone wondering whether they were 'not working,' to get away for coffee, and so on—all of which meant they heard more news from the other executives they talked with [p. 47].

Burns' study (1954) also appears to support the notion of greater communication flexibility for managers in staff positions.

Summary

Since the studies that have compared the attitudes and behavior of employees in line and staff positions are all relatively recent, it is impossible to determine if the amount of difference between the two types of hierarchies has been decreasing over the long term, as some writers have suggested. The validity of this contention probably will not be established until we have strictly comparative data on line-staff differences from different time periods. In the meantime, however, the available evidence indicates that differences in conditions between the two types of positions are still large enough to produce significant differences in both attitudes and behavior. Staff managers derive less satisfaction from their jobs, feel they have to be more other-directed, and exhibit different patterns of behavior. The only area that failed to produce consistent differences was the degree of importance attached to various job-related factors and needs. Apparently this type of attitude was little affected by organizational conditions surrounding line and staff jobs.

The evidence on the attitude and behavior differences between managers in line and staff positions presents a picture of several factors that may be responsible for line and staff conflicts. The staff manager's position, for instance, requires him to be knowledgeable about his specialty and spend considerable time with a variety of types of executives; yet it provides him with little opportunity to use his knowledge for decision making unless he usurps the authority of the line. The line manager's position, on the other hand, places him in a situation in which he is expected to take advice, on decisions for which he has responsibility, from a man who frequently is younger and has been with the company a shorter period of time and who is not supposed to know the whole picture as he is. This is not a situation that lends itself to harmonious line-staff relations. The fact that managers in line and staff hierarchies see different traits being required for success and report different levels of satisfaction suggests that one way of easing some of the potential line-staff conflicts may be to take into account the psychological demands of the two different types of jobs in the selection and placement of managers.

SPAN OF CONTROL

Span of control is defined as the number of subordinates a manager is responsible for supervising. Since the early days of classical organization theory (Urwick, 1935), writers have prescribed what they felt to be an optimum span of control. In most instances, the classical theorists have assumed that a small span of control (few subordinates) is a "good thing" for any organization. Typical prescriptions have recommended spans of control between three and six (Dale, 1952). However, classical organization theorists generally have failed to consider what factors might call for a change in the size of the span of control. Recognizing this, Fisch (1963) has recently argued that such factors as level of management, company product diversification, and the personalities of the managers should be considered in determining the optimum span of control.

A vigorous dissent from the classical approach to the span of control was made by Worthy (1950). His radical suggestion was that a large span of control is good since it provides better communication and greater opportunities for individual growth and initiative. With this statement by Worthy, the battle line between the two approaches to organization theory was clearly drawn. However, researchers have not been quick to join in the fray, and as late as 1960 Pfiffner and Sherwood stated: "There has been relatively little, if any, empirical study of span of control [p. 156]." The majority of the research that has been done on the topic has been concerned with discovering variables that are related to the size of the span of control found in existing organizations. Entwisle and Walton

(1961), for example, found that the size of the organization and the function of the organization were both related to existing spans of control. This type of research undoubtedly represents a necessary first step in determining which conditions favor a large span of control and which favor a small span of control, but it does not provide data on relationships to job attitudes or performance. In fact, only one such study was located in the literature.

Attitudes

No empirical studies were found that attempted to determine the relationship between span of control and employee job attitudes. Worthy (1950) did report that in one company a large span of control had been found to be positively related to high morale; however, he presented no published data to support this point, and thus it is difficult to evaluate his claims.

Behavior

A study by Woodward (1958) found a relationship between the size of the span of control at the first level of supervision and company performance when several English companies were viewed. The companies were divided into three groups based upon the method of production used—unit (e.g., production of unique units to customers' orders), mass (e.g., production of large numbers of identical units by assembly-line techniques), and process (e.g., production of chemicals, etc., by continuous processing). For each of the three types of production, the most successful companies were those that were near the median span of control for that production type. These median spans of control (23 for mass, 49 for process, and 13 for unit) for the successful companies were substantially larger than those suggested as optimum by the classical theorists. Woodward's study thus gives at least tentative support to Worthy's point that a large span of control can produce high performance. However, it is important to remember that Woodward's study was concerned with the span of control of first-line supervisors, and although a relatively large span of control may have proved to be optimum at this level, there is no reason to believe the same necessarily will be true of the upper levels of management. Perhaps the most significant point about the Woodward study is that the method of production was found to be an important variable in determining which span of control at the first level of supervision was optimum for producing high performance.

Summary

The lack of research makes it impossible to state that any particular span of control is best for producing high performance or positive job attitudes. What data there are suggest that it is not reasonable to expect any one span of control to be ideal for all organizational situations. Level of management and type of production appear to be two variables that influence the optimum span of control for a given situation, and undoubtedly further research will uncover other variables that are important.

SIZE: SUBUNITS

For the purposes of this review, any grouping of the members of a business organization that systematically excludes part of the membership of that organization is considered an organization subunit. Primary work groups, departments, and factories (in multifactory companies) have been frequently studied organization subunits. Previous summaries of the research evidence on the relationship between organization subunit size and job attitudes and performance typically have concluded that small organization subunits are superior on all counts (e.g., Strauss & Sayles, 1960; Viteles, 1953). Morale, absence rates, turnover rates, accident rates, and productivity are all considered to be better in small than in large organization subunits. An example of the broad statements supporting small organization subunits is Viteles' conclusion that "The size of the work group affects output and attitudes, which both tend to be better in smaller sized groups [p. 146]."

Recently several articles (Herbst, 1957; Indik & Seashore, 1961) have contained theories attempting to explain the beneficial effects of small organization subunits. Undoubtedly the appearance of these theories is due to the aforementioned widespread acceptance of the belief that the experimental evidence con-

TABLE 2

STUDIES OF RELATIONSHIPS BETWEEN ORGANIZATION SUBUNIT SIZE AND JOB ATTITUDES AND JOB BEHAVIOR

Attitude or behavior studied	Investigators	Type of subunit considered	Relationship found[a]
Job satisfaction	Talacchi (1960)	Factories	Negative
	Kerr, Koppelmeier, & Sullivan (1951)	Departments	Negative
	Indik & Seashore (1961)	Departments	Negative
	Katzell, Barrett, & Parker (1961)	Departments	Negative
	Campbell (1952)	Work groups	Negative
	Worthy (1950)	?	Negative
	Indik & Seashore (1961)	Automobile dealerships	Zero
Absenteeism	Revans (1958)	Gas works	Positive
	Revens (1958)	Factories	Positive
	Revans (1958)	Factories	Positive
	Acton Society Trust (1953)	Factory	Positive
	Baumgartel & Sobol (1959)	"Plants" (airline locations)	Positive
	Research Council for Economic Security (Baumgartel & Sobol, 1959)	Plants	Positive (?)
	Hewitt & Parfitt (1953)	Departments	Positive
	Indik & Seahore (1961)	Departments	Positive
	Kerr, Koppelmeier, & Sullivan (1951)	Departments	Positive
	Metzner & Mann (1953)	Work groups (blue-collar)	Positive
	Argyle, Gardner, & Cioffi (1958)	Work groups	Curvilinear
	Metzner & Mann (1953)	Work groups (white-collar)	Zero
Turnover	Indik & Seashore (1961)	Automobile dealerships	Positive
	Kerr, Koppelmeier, & Sullivan (1951)	Departments	Positive
	Mandell (1956)	Departments	Positive
	Argyle, Gardner, & Cioffi (1958)	Work groups	Zero
Accidents	Revans (1958)	Mines (Britain)	Positive
	Revans (1958)	Factories (Britain)	Positive
	Revans (1958)	Departments (Asia)	Positive
	Revans (1958)	Mines (U. S. A.)	Curvilinear
	U. S. Department of Labor (Revans, 1958)	Factories (U. S. A.)	Curvilinear
	National Safety Council (Revans, 1958)	Factories (U. S. A.)	Negative
Labor disputes	Cleland (1955)	Factories	Positive
	Revans (1958)	Mines	Positive
Productivity	Katzell, Barrett, & Parker (1961)	Company divisions	Negative
	Indik & Seashore (1961)	Departments	Negative
	Marriott (1949)	Work groups	Negative
	Revans (1958)	Mines	Curvilinear
	Revans (1958)	Retail stores	Curvilinear
	Herbst (1957)	Retail stores	Curvilinear
	Indik & Seashore (1961)	Automobile dealerships	Zero
	Argyle, Gardner, & Cioffi (1958)	Work groups	Positive

[a] A positive relationship indicates a trend for the attitude or behavior to become more frequent as size increases. A negative relationship indicates a trend for the attitude or behavior to become less frequent as size increases. A curvilinear relationship indicates a trend for the middle-sized subunit to exhibit the greatest or the lowest frequency of the attitude or behavior.

clusively supports the proposition that smaller organization subunits are always beneficial. Virtually ignored in the rush to accept this simple picture of the effects of subunit size and completely ignored in theoretical attempts to explain this picture has been the effect of two other variables: job level within the organization subunit, and type of organization subunit (e.g., work group, department, factory). These variables may well function to limit the degree to which it is possible to generalize about the beneficial effects of small organization subunits. For example, increasing the size of a department may raise the absence rate among the department's production workers, but it is doubtful that it will also increase the absence rate of the department's management force. Likewise, increasing the size of a primary work group may cause lower morale in that work group, but increasing the size of a factory by adding more small work groups may very well not decrease the morale of the total plant. Thus, before we accept the conclusion that all types of large subunits, regardless of job level being considered, are always inferior to small ones, it may be well to see if the evidence justifies such a conclusion.

Since there are a relatively large number of studies that have investigated the relationship of subunit size to both attitudes and behavior, Table 2 was constructed to summarize most (but not all) of these studies. This table will serve as the basis for most of the subsequent discussion in this section on subunit size.

Attitudes

Job Satisfaction and Morale. As Table 2 indicates, there is little doubt that subunit size is significantly related to differences in job attitudes. The evidence is strong that workers in small departments and work groups are better satisfied than workers in large departments or groups, as six of the seven studies reviewed show results in this direction. However, the evidence does not yet support the sweeping conclusion that all types of employees in smaller organization subunits are more satisfied than those in larger subunits, since no studies were located, for example, that compared managerial job attitudes in different size organization subunits. Thus, it is impossible to reach any conclusion about the relationship between subunit size and the job satisfaction of managers.

Talacchi (1960) studied the level of job satisfaction among employees in 41 plants that were subunits of five larger organizations. His results showed a strong trend (r's ranging from $-.42$ to $-.81$) for job satisfaction to decrease as plant size increased. Kerr, Koppelmeier, and Sullivan (1951) found that for 894 workers in 29 departments in two electronics plants job satisfaction was correlated $-.46$ with size of department. Indik and Seashore (1961) measured the amount of intrinsic job satisfaction expressed by workers in 32 package delivery departments, varying in size from 15 to 61, and found that increased size led to lower satisfaction. Likewise, Katzell, Barrett, and Parker (1961) found a significant trend for workers in large warehouses to express lower job satisfaction than workers in small warehouses (with 23 out of 23 correlations being in this direction). Campbell (1952), in studying workers' satisfaction with incentive pay plans as a function of the size (from under 20 to over 100) of the work group to which they belonged, found that workers in the smaller groups felt they had better knowledge of how their pay plan operated and were more satisfied with their pay. Finally, Worthy (1950) reported: "Our researches demonstrate that mere size is unquestionably one of the most important factors in determining the quality of employee relationships: the smaller the unit the higher the morale, and vice versa [pp. 172–173]." Regrettably, Worthy failed to document this strong statement with any published data on the size of the units studied, the type of employees studied, or the measure of morale used.

The only study that failed to show a negative relationship between subunit size and job satisfaction was one reported in the previously mentioned publication by Indik and Seashore (1961) which concerned employee job satisfaction in automobile dealerships. Since these dealerships were relatively autonomous, and perhaps not completely accurately classified as subunits, the failure to find a strong negative relationship in this particular instance is probably not too significant.

Other Attitudes. Three studies (not shown in Table 2) investigated the relationship between organization subunit size and job attitudes other than those concerned with satisfaction. Hemphill (1950) found that as group size increased, members expressed attitudes of greater tolerance for leader-centered direction of group activities. In addition, Hemphill (1956) found that small groups were better characterized by attitudes of high intimacy, high hedonic tone, and low control than were large groups. It should be noted that both studies by Hemphill employed subjects from very diverse types of groups (ranging from industrial work groups to members of a sorority), and thus it is difficult to assess their relevance for industrial organizations.

Seashore (1954) studied the relationship between group cohesiveness and size of work group for more than 200 groups in a large industrial organization. The results showed a general trend for feelings of cohesiveness to decrease with increasing subunit size. However, there was also a tendency for the smallest size groups to fall at the two extremes of cohesiveness; that is, they were either extremely high or extremely low in cohesiveness. Nevertheless, Seashore's general conclusion was that group size is negatively related to group cohesiveness attitudes.

Behavior

Absenteeism. The greatest amount of evidence demonstrating a relationship between organization subunit size and job behavior appears on the topic of absence rates. As shown in Table 2, 10 of the 12 studies reviewed found positive linear relationships between absence rates and the size of subunits. This finding appeared for factories, departments, and primary work groups. Again, however, the finding is limited to absence rates among employees at the blue-collar level, since the only study that separately looked at white-collar workers' absences found no relationship between subunit size and absence rates.

Revans (1958) has provided an excellent review of much of the data on the topic of subunit size in relation to absenteeism. He reported that in five randomly chosen gas works,[2] ranging in size from 67 to 3,430 employees, the duration of absence due to accidents and the log of the size of the works were correlated +.91. In addition, size of works was correlated +.62 with absence rates due to sickness and other factors. Revans also reported two further studies dealing with size of factory in multifactory organizations. Each study found significant positive correlations between log size of factory and absence rates. A study by the Acton Society Trust (1953) reported a correlation of +.44 between size of factory (ranging from under 100 employees to more than 1,000) and absence rates. Baumgartel and Sobol (1959) did a well-controlled study on the relationships between the size of work locations or "plants" of an airline and absenteeism and found a strong tendency for absenteeism to increase with size of the work location. These investigators also cited the report on Prolonged Illness Absenteeism rates. And Metzner and Mann for Economic Security, which showed that smaller-sized plants had lower prolonged illness rates than medium-sized plants. However, the same report indicated that there was not a consistent linear relationship between plant size and illness. Several other studies, Hewitt and Parfitt (1953), Indik and Seashore (1961), and Kerr, Koppelmeier, and Sullivan (1951), have all found that larger departments of a company have higher absenteeism rates. And Metzner and Mann (1953) found that among blue-collar workers small work groups (seven and under) had fewer absences than did large work groups (over seven).

Only 2 of the 12 studies dealing with subunit size and absenteeism failed to show positive relationships between these two variables. Argyle, Gardner, and Cioffi (1958) reported that they found a curvilinear relationship between absence rates and work-group size (ranging from 1 to 20+), with the lowest absence rates occurring in the middle-sized groups. Metzner and Mann, in their above-cited study (1953), found an insignificant relationship, although in the expected direction, between absence rates and work-group size for white collar workers. (It should also be noted that in the previously cited Baumgartel and Sobol study of airline locations, the effects of location size were less for white-collar workers than for blue-collar workers.)

---

[2] Much of the literature on the relationship between behavior and size fails to point out whether the units studied (e.g., gas works, factories, or mines) are parts of larger organizations or are separate corporations. For the purposes of this review, it was decided that unless the author specified he was dealing with separate companies or firms, all data on factories, gas works, and mines would be dealt with as data on organization subunits.

Turnover. Table 2 shows that three of four studies found that turnover is greater in large units compared to small units. However, none of the studies compared the turnover rates of different-sized factories, and none of the studies considered turnover rates among managerial personnel. Indik and Seashore (1961) found that more employees were terminated in large than in small automobile dealerships. Kerr, Koppelmeier, and Sullivan (1951) found a significant relationship between turnover and department size ($r = +.49$). Likewise, Mandell (1956), in a study of 320 companies, found that large offices tended to have higher turnover rates than did small offices. On the other hand, Argyle, Gardner, and Cioffi (1958) did not find any relationship between subunit size and turnover.

Accident Rates. Revans (1958) has thoroughly reviewed the literature on the relationship between accident rates and organization subunit size. As the six sets of data covered in his review indicate, there is not a consistent pattern of relationship between these two variables. In fact, three different patterns were found: positive, curvilinear, and negative relationships. A review of the evidence cited by Revans leads to the conclusion that accident rates may be dependent upon the type of industry considered and upon the technological development of the country in which a study is carried out. At the present time it seems wise to speak only of the type of relationship that exists for a given industry in a given country, and not for industry in general. For example, as shown in Table 2, the evidence is fairly convincing that for United States industry large size is not associated with high accident rates, while for British industry large size does appear to be associated with increased accident rates.

Revans' analysis of the accident rate for British coal miners and quarriers in 1950 showed a strong positive relationship between mine size and number of accidents. Likewise, the Annual Report of the Chief Inspector of Factories for 1956 (cited by Revans) showed that the compensable accident rate rises steadily with the average size of factory for British industrial firms. Also reported by Revans was a strong positive correlation (+.47) between accident rates and department size for steel companies operating in Asia.

Turning to the data for American industry, Revans cited three studies that indicated either a curvilinear or negative relationship between subunit size and accident rates. The data for American mines, in contrast to similar data for mines in Britain, showed the highest accident rates occurring in middle-sized mines employing 100 to 300 workers. Data for United States factories were provided in two studies which showed somewhat different results from each other. Bulletin 1164 of the United States Department of Labor (cited by Revans) on "Work Injuries in the United States during 1952" points to a curvilinear relationship in the following quotation:

> The larger establishments, which can afford trained safety engineers and which conduct intensive safety programs, generally have the lowest rates. Usually the medium-size plants have the highest rates, and the smallest establishments show rates somewhat below the medium-size plants but above the average for the industry [quoted by Revans, 1958, p. 209].

In a report from the National Safety Council on 3,500 United States manufacturing plants, Revans cited data showing a strongly declining accident rate with increasing plant size.

Labor Disputes. Data contained in two sources (Cleland, 1955; Revans, 1958) show a positive relationship between subunit size and labor disputes. Cleland's data demonstrated that as plant size increases, the probability of an industrial strike rises significantly. Similarly, Revans noted that British coal miners in larger mines were more sensitive to nonmonetary issues (although not to monetary issues) than were miners in smaller locations. On the basis of these two reports it is tempting to conclude that small subunits are more likely to have fewer labor problems than larger subunits. However, one must be cautious in making this interpretation, since it is not size of operation, as such, that affects the number of labor disputes, but rather the type of technology that is associated with size (see Summary at the end of this section).

Productivity. As Table 2 shows, the studies on the relationship between performance and organization subunit size do not present a clear-cut picture. Thus, the available evidence does not support the widely believed assumption that small subunits are superior to large ones in terms of job performance.

Three studies found better performance in smaller subunits than in larger units. For example, Katzell, Barrett, and Parker (1961) found that product value, productivity, and profitability were all lower in the large divisions of the company they studied. Indik and Seashore (1961) found a negative correlation between the size of delivery organizations and rated overall performance of these units. Marriott (1949), in a widely cited study of the relationship between work-group size and production in two automobile factories, concluded from his data that: "Low, but significant correlations were obtained which demonstrate an inverse relationship between output and size, the smaller sized groups showing consistently larger output in each factory [p. 56]."

In contrast to the results from the above three studies, were the findings from five other studies and reports; three of them found curvilinear relationships between subunit size and production, one found essentially no relationship, and one a positive relationship. Revans (1958) reported that data gathered by the National Coal Board in Britain for Years 1948–53 showed a curvilinear relationship between mine size and output per miner, with the highest output being obtained in middle-sized mines where 1,500 to 2,000 men were employed. Revans (1958) and Herbst (1957) both present data on sales of retail sales units, and both report curvilinear relationships to size, with the middle-sized units again performing best. Indik and Seashore (1961), on the other hand, reported a lack of significant relationship between size of automobile dealerships and job performance, and Argyle, Gardner, and Cioffi (1958) found a slight tendency for larger work groups to achieve higher productivity than smaller groups.

Summary

The literature on subunit size shows that when blue-collar workers are considered, small size subunits are characterized by higher job satisfaction, lower absence rates, lower turnover rates, and fewer labor disputes. The evidence does not show, however, a consistent relationship for blue-collar workers between accident rates and subunit size, or between productivity and subunit size. Furthermore, none of the studies reviewed investigated the relationship between managers' job attitudes and job behavior and organization subunit size.

Future research might well consider the problem of assessing the relative impact on attitudes and behavior of the size of the different organization subunits of which an individual is a member. A typical worker is a member of at least three organization subunits: a primary work group, a department, and a factory or office. Although the evidence indicates that large-sized subunits are associated, for example, with high absenteeism, it is impossible to know on the basis of past studies if it is large size in the primary work group, department, or factory that is crucial in this relationship. Previous investigators have failed to control for variation in the size of the other types of subunits of which an individual is a member while they are studying the effects of size of one type of unit. For example, no study has compared absenteeism among work groups that are part of the same size departments and factories. Similarly, we do not know if factory size is related to absenteeism when all factories considered have the same size departments and work groups. It may well be that absenteeism is determined by work-group size but because large work groups are found in large factories and departments a spurious relationship exists between absenteeism and department size and between absenteeism and factory size.

At the present time, the weight of the evidence clearly suggests that small-sized subunits are desirable for a number of reasons. Although small-sized units may not improve

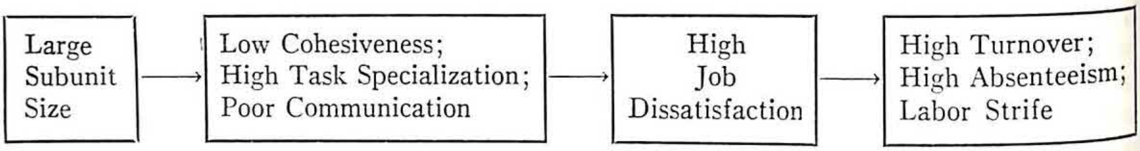

Large Subunit Size
Low Cohesiveness; High Task Specialization; Poor Communication
High Job Dissatisfaction
High Turnover; High Absenteeism; Labor Strife

FIG. 1. Hypothesized effects of subunit size.

job performance directly, the fact that they are associated with low turnover and low absenteeism argues strongly for their value. The question then arises as to why small subunits seem to have certain advantages over large units in affecting attitudes and behavior. Figure 1 presents a diagram that offers a possible explanatory model for the effects of organization subunit size on satisfaction, turnover, absenteeism, and labor strife. As subunits increase in size, it is probably increasingly difficult to maintain high cohesiveness and good communications. Further, task specialization may be more prevalent. The factors of low cohesiveness, high task specialization, and increased communication difficulties might tend in turn to lead to high job dissatisfaction. As Brayfield and Crockett (1955), among others, have pointed out, there is good reason to believe that dissatisfaction leads to higher turnover and absenteeism rates. Also, labor strife logically would seem to be affected by increases in job dissatisfaction.

On the basis of the diagram presented in Figure 1, there is no necessary reason for large units to lead directly to decreases in productivity. Although the model does predict that large units will be characterized by, among other things, higher job dissatisfaction, reviews of the literature have not found that job dissatisfaction is associated with low productivity. Nor is there evidence that task specialization, low cohesiveness, and communications problems must necessarily lead to lower productivity. It is not surprising, therefore, that the evidence on the relationship between productivity and subunit size is inconsistent.

SIZE: TOTAL ORGANIZATIONS

All of the studies discussed in the previous section were concerned with the relationships between size of some type of organizational subunit and attitude and behavioral indices. In the present section we will review only those studies that have dealt specifically with the total organization as the unit of size. It will become evident that the number of such studies on total organization size is much lower than the number of studies on subunit size.[3]

The distinction between subunit size and total-organization size has seldom been made clear either in previous empirical studies dealing with size or in various interpretations of these studies presented in textbooks concerned with organizational behavior. It is conceivable, for example, that although working in a large subunit has disadvantages (as demonstrated by the evidence in the preceding section), working in a large total organization might have advantages as long as the subunits within the organization are relatively small. Therefore, it seems necessary to keep the subunit/total-organization distinction clearly in focus because the effects of one type of size (e.g., total organizational size) may be confounded by the effects of the other type of size (i.e., size of subunits within total organizations).

The absence, in most previous studies of size, of a distinction between subunits and total organizations has been compounded by the failure to take into account the job level of the employees. As was pointed out in the previous section, organizational level may interact with size such that conclusions pertaining to lower-level employees may not hold up when applied to higher-level personnel.

---

[3] By the term "total organization" we mean a total operating company headed by an executive with the title "President." It is admittedly difficult at times to determine whether a "company," in the loose sense of the word, should be considered a separate total organization, in our terms, or merely a subunit of an even larger "corporation." In general, if a company has a chief executive with the title of president and if that company can sell stock independently of other "companies" all under the same corporate holding entity, we would consider it a total organization.

This possible interaction effect of level could presumably influence results both for studies of subunit size and for studies of total-organization size. With the foregoing considerations in mind, we now turn to a review of the few studies that have been concerned with total-organization size.

Attitudes

At the rank-and-file worker level, two articles have appeared in the literature that *seem* to be concerned with the relationship of total-organization size (rather than subunit size) to job attitudes. The first of these two articles was by Benge (1944). Apparently sampling from a number of different companies and using the "composite attitude of employees in each company toward the boss" as an index of morale, he found that "morale of employees of small companies is appreciably better than that in larger companies [p. 104]." These results are extremely difficult to evaluate, however, since neither the number or type of respondents nor the number of companies on which the results were based was specified in the article.

The second study that seems to be concerned with the relationship of total-organization size to job attitudes of rank-and-file workers is one reported by Talacchi (1960). The study is described as an investigation involving 93 "organizations"; however, as one of the tables (Table 2 in the article, p. 411) indicates, at least 45 of the 93 organizations are in reality "plants" that comprise parts of only five companies. Thus, most of the findings reported in this article are difficult to interpret if one is interested in the effects of total-organization size, since the distinction between subunits and total organizations has been completely blurred. Nevertheless, it appears that the article does present some evidence bearing on the effects of size of the total organization on employee job satisfaction. The key result reported in Talacchi's article was a correlation of −.67 between organization size and the overall index of employee satisfaction as measured by the SRA Employee Inventory. Employee dissatisfaction was, therefore, apparently greater within large organizations (which ranged up to a maximum size of 1,800 employees). Aside from the limitations in the study that have already been pointed out, this finding must be interpreted cautiously because it is possible that the larger organizations contained larger subunit work groups. In other words, it is possible that what seemed to be an effect of organization or, at least, plant size may have been due to work-group size.

At the management level of organizations, several recent articles by Porter (1963c, 1963d, 1964) examined the relationship between three types of job attitudes and the size of company for which each manager worked.[4] In this questionnaire study of more than 1,500 managers, respondents were placed into one of three size-of-company categories: small (companies of under 500 employees), medium (500–4,900 employees), and large (more than 5,000 employees). One part of the data dealt with need satisfactions and indicated that the effects of size were modified by the level of management being considered. At the lower and lower-middle management levels, managers from smaller companies were more satisfied than those from larger companies. At the upper-middle and vice-president levels, the reverse was true, with the managers from the larger companies indicating greater satisfaction. With respect to a different area of perceptions, when respondents were asked to rank 10 personality traits in terms of their importance for job success, it was found that managers from larger companies placed more emphasis than did managers from smaller companies on innerdirected or non-organization-man traits compared to otherdirected or organization-man traits. When respondents (across the intermediate three levels of management) were asked to use semantic differential scales to describe their jobs, managers from larger companies were slightly more likely to apply the terms "challenging," "interesting," and "competitive" to their jobs than were small company managers. (For both the personality trait and job-description results, management level did not seem to affect the findings. In other words, the effects of size were fairly consistent from one management level to the next.) These latter findings were interpreted as indicating that large companies place no more emphasis, perhaps even less, on conforming, organization-man type of behavior than do smaller companies.

---

[4] It should be pointed out that Porter also did not control for size of work groups or subunits in which an individual manager worked within his company. However, it is assumed that the concept of work group—number of individuals reporting to a single supervisor—is a less meaningful and important variable at managerial levels because of the nature of the work situation. Hence, it is probably not as necessary to control its effects in assessing the effects of total size when the employees being considered are managers. Nevertheless, it is possible that this variable did have some undetermined effect on Porter's findings.

Behavior

With the exception of one quite limited area of investigation, there are almost no studies that specifically compare large total organizations or companies to small total organizations on behavioral measures other than attitude data.[5] Undoubtedly the major reason for this lies in the difficulty of obtaining a sample of organizations (as compared to a sample of individuals or a sample of subunits within organizations) in sufficient quantity and with sufficient variation in total size.

The one type of behavior that has been investigated in relation to total-organization size concerns the rate of managerial succession into top-level jobs. The first such study, by Grusky (1961), utilized *Fortune's* annual list of the 500 largest United States companies to select two groups of organizations differing on total size. His results indicated that the frequency or rate of succession was positively related to size such that there was more rapid turnover in the uppermost management positions within larger-sized companies. A study by Kriesberg (1962) seemed to confirm Grusky's findings. However, a later study by Gordon and Becker (1964), which involved a reanalysis of the data used by Grusky, showed very little relationship between size and rate of managerial succession. Thus, the possible effects of size on this particular variable of organizational behavior remain obscure at this point.

Summary

A review of relevant journals shows that very few studies have been carried out on the relationship of size of total organization to either job attitudes or job behavior. Two studies of worker attitudes that seem to fit in this category of investigations indicated that job satisfaction and morale were lower in larger organizations. However, both studies were reported in such a way that it is difficult to draw accurate or meaningful conclusions concerning the effects of organization size. At the managerial level, Porter (1963c, 1963d, 1964) found no overall advantage for either large or small organizations in relation to need satisfactions; he did find, however, that managerial jobs in larger companies were seen as requiring a somewhat greater emphasis on inner-directed behavior and as having a slightly greater amount of challenge and interest when compared to similar managerial jobs in smaller companies.

Overall, the findings relating total-organization size to job attitudes do not present as clear a picture as is the case for findings dealing with subunit size. It is entirely possible that the negative effects of large subunit size on job attitudes may not extend to large total-organization size. As has been suggested,

> An increase in the total size of an organization—with the consequent technological advantages of large-scale operation—will not necessarily reduce the morale and job satisfaction of employees as long as intra-organization work units are kept small [Porter, 1963d, p. 61].

This is a hypothesis that could be fairly easily confirmed or disconfirmed by future studies. Further research on total-organization size should also take into account the possibility that size of the total units may have one kind of effect on attitudes at the rank-

---

[5] The previously discussed article by Talacchi (1960) has sometimes been cited as an article that presents data on the relationship of "organization" size to job behavior, such as turnover rates and absenteeism rates, but no such data are given in the article. Instead, the findings are presented in terms of the relationship of size to satisfaction (as already noted above) and the relationship of satisfaction to turnover and absenteeism. Not only are there no data presented on the direct relationship of size to turnover and absenteeism, but also the data that are presented concerning these latter two variables involve intraorganization rather than interorganization comparisons. Thus, this article does not provide behavioral data in relation to size of total organization.

and-file level, but a reversed effect at managerial levels, especially the higher levels. Again, this possibility would not be too difficult to test with additional research.

In the area of behavior, there is almost a complete lack of information on the effects of total-organization size. It would seem that future studies on size could, among other things, attempt to obtain measures of absenteeism, turnover, and similar rates in relation to total-unit size just as past studies have obtained this kind of data in relation to subunit size. At the moment, at least in the psychological literature, we do not even know whether the United States Steels, the Standard Oils, the General Motors, etc., have higher or lower turnover or absenteeism rates throughout their total organizations than do the local Acme or Ace manufacturing companies. (Again, it must be emphasized that if such comparisons of total organizations are to be meaningful, the effects of differences in the sizes of various subunits among the organizations must be partialed out and job level controlled.) Perhaps total-organization size will turn out to be an irrelevant variable in affecting such behavioral indices. At any rate, it would seem worthwhile to know whether this is the case or not.

SHAPE: TALL OR FLAT

Tall and flat organization structures are generally distinguished on the basis of the number of levels in the organization relative to the total size of the organization. A flat organization structure is one where there are few levels relative to the total size of the organization and a tall organization structure is one where there are many levels relative to the total size of the organization. Another way of stating this is to say that the degree to which a structure is tall or flat is determined by the average span of control within the organization.

Attention has been focused on the relative merits of these two types of structures since 1950, when Worthy published his widely cited article on "Organizational Structure and Employe Morale" (see Gardner & Moore, 1955; Viteles, 1953; Whyte, 1961). Worthy's (1950) basic conclusion was that: "Flatter, less complex structures, with a maximum of administrative decentralization, tend to create a potential for improved attitudes, more effective supervision, and greater individual responsibility and initiative among employes [p. 179]." Although Worthy published no empirical evidence to support his statements, and although his observations were based upon his experiences in a single company, his opinions are frequently cited by other authors to support their contention that flat organization structures produce better performance and better job attitudes than do tall ones. Worthy's opinions represent direct contradiction to those put forth by the classical organization theorists (e.g., Graicunas, 1937) who contend that a tall organization structure improves performance by allowing for close supervision and therefore complete understanding by the supervisor of the subordinate's activities.

Despite the fact that this significant disagreement about the relative advantages of tall and flat organization structures has provoked considerable comment in the literature, the topic did not generate any empirical research until 1962.

Attitudes

Three studies have compared the job satisfactions of employees in tall organizations with those in flat organizations. Meltzer and Salter (1962) published a study that reported on the job satisfactions of 704 physiologists (in nonuniversity organizations). They classified their questionnaire respondents by size of company (fewer than 20 professional employees, 21–50, and 51 or more), and by number of levels of administration within the organization (1–3 levels, 4–5, and 6 or more), and found generally insignificant relationships between tallness or flatness and job satisfaction. Their results, therefore, did not confirm Worthy's theory about flat structure producing better job attitudes. However, it should be noted that Meltzer and Salter studied research organizations of extremely small size, and hence their results may have limited generality.

A study by Porter and Lawler (1964) investigated the relationship between the need satisfactions of over 1,500 managers and the type of organization structure in which they worked. As in the Meltzer and Salter study, managers were classified as working in either a tall or flat organization structure based upon the number of levels in the organization relative to its total size. The results showed no clear overall superiority of flat over tall organizations in producing greater need satisfaction among managers. However, two qualifications to this general finding were noted: First, organization size seemed to have some effect on the relationship between type of structure and the degree of need satisfaction. In companies employing fewer than 5,000 people, managerial satisfaction was greater in flat rather than in tall organizations. For companies with 5,000 employees and over, the picture was reversed with a tall type of structure producing perceptions of greater need satisfaction. The second qualification was that the effects of organization structure on need satisfaction appeared to vary with the type of psychological need being considered. A tall type of structure was associated with greater satisfactions in the security and social need areas, whereas a flat structure was associated with greater satisfaction in the self-actualization need area.

Porter and Siegel (1964) have replicated the Porter and Lawler study with one essential difference—the sample. The Porter and Siegel sample of close to 3,000 managers was an international sample representing middle and upper levels of management in a wide variety of sizes and types of companies in 13 countries. The results of the newer study generally agreed with those found by Porter and Lawler. Porter and Siegel found that for companies of less than 5,000 employees, flat structures produced greater managerial need satisfaction, whereas for companies of 5,000 and over, there was no difference between tall and flat structures in producing managerial need satisfactions.

Behavior

Meltzer and Salter's (1962) study was the only empirical investigation that compared the job performances of individuals working in organizations with tall and flat structures. They found generally insignificant relationships between tallness and flatness and job performance (as measured by the publication of research articles). Only in large-size organizations did they find a significant trend and there it was in the direction of greater productivity in tall than in flat organizations.

Summary

The evidence does not support Worthy's (1950) sweeping generalization that a flat organization structure produces greater job satisfaction and improved job performance. The evidence points to organization size as one of the factors affecting the relative advantages of tall and flat organization structures. Two of the studies reviewed found that in relatively small organizations a flat organization did appear to be advantageous in terms of producing managerial job satisfactions. However, for relatively large organizations one study found that tall organization structures produced greater job satisfaction, and one study found that tall organization structures fostered greater productivity. Thus, it appears that the advantages of a flat structure not only decrease with increasing organization size, but that in relatively large organizations a flat structure may sometimes even be a liability.

One explanation for the tendency for the advantages of a flat organization structure to decrease with increasing organization size can be found in the characteristics generally attributed to tall and flat organization structures. Advocates of flat organization structures claim that because of the large average span of control subordinates will have greater freedom and autonomy to make decisions. As a result of this relatively greater freedom and autonomy, individuals are supposed to contribute more to the organization and receive greater satisfaction from their jobs. On the other hand, a tall structure increases supervisory controls and allows superiors to better coordinate the activities of their subordinates. In a small organization problems of coordination and communication do not tend to be severe, simply because the organization is small. Thus, in a small organization there would be little advantage in a tall structure and, in fact, since it tends to amplify the disadvantages associated with tight managerial control, a tall structure probably is a liability in a typical small organization. In large organizations, on the other hand, problems of coordination and communication are complex. Thus, for large organizations a taller type of structure may be needed to overcome these problems and allow managers to supervise their subordinates more effectively.

Future research will undoubtedly show that other variables such as type of company—for example, retail trade versus manufacturing firms—will have an important bearing on which type of structure is more advantageous in terms of job attitudes and job performance. Worthy (1959) himself has pointed out that different degrees of flatness may be desirable depending upon the type of organization considered.

SHAPE: CENTRALIZED OR DECENTRALIZED

> The contemporary trend in large-scale organization is toward decentralization, but we must realize that decentralization is several things to different people. There are those who view it entirely in terms of decision-making; others see it from the standpoint of geographical dispersion of plants and installations; and still others approach it as a philosophy of corporate life, a set of organization values with sociological, psychological, and spiritual facets [Pfiffner & Sherwood, 1960, p. 190].

There is no reason to believe that these different types of decentralization produce similar effects on job attitudes and performance. Thus, the first point that should be considered about any study concerned with the effects of decentralization is the measure of decentralization that is used.

According to Chandler (1956), the movement to decentralization can be dated to the 1920s and the management policies of Alfred P. Sloan, Jr., at General Motors. That the topic of decentralization is currently in the air is clearly demonstrated by the great number of recent articles that have dogmatically set forth plans for a decentralized corporate way of life (see Cordiner, 1956; Ginzberg & Reilley, 1957; Lawrence, 1958; Smith, 1958). Most writers have claimed that by offering increased autonomy for the individual regardless of his level within the company, decentralization improves both job attitudes and performance. Since the topic of decentralization has received considerable attention for the last forty years, it is not surprising that there has been a certain amount of research on the topic. However, most of the research on the effects of decentralization has been of the case study variety (e.g., Drucker, 1946; Given, 1949; Selznick, 1949), and therefore not within the scope of this review. In fact, only four studies were found that compared the job attitudes and behavior of employees in centralized and decentralized companies.

Attitudes

A study by Baker and France (1954) compared the attitudes of managers in centralized and decentralized industrial relations departments. Their classification of centralized or decentralized departments was based upon the level at which decisions relative to industrial relations were made. When they asked managers which type of structure produced the best intramanagement relations, managers who worked in companies with centralization favored centralization, and managers who worked in companies with decentralization favored decentralization. Baker and France further found that companies with decentralized industrial relations departments revealed no greater satisfaction among their plant managers than did those companies with centralized industrial relations departments. Although this latter finding does not support the claims of those who favor decentralization, it is important to remember that the study was concerned with decentralization in only one type of department, and the attitudes of only one level of management. Thus, it is impossible to know if the findings of Baker and France are applicable to the effects of decentralization on a company-wide basis.

Litzinger (1963) has compared the attitudes of bank managers who were under centralized management with those who were under decentralized management. His results indicated no clear attitude differences between the centralized and the decentralized groups of managers. However, the results are difficult to interpret because it is not clear from the article what measure of decentralization was used.

Behavior

Two studies were found that compared the behavior of employees in centralized companies with the behavior of employees in decentralized companies. Carlson (1951) found that executives in decentralized companies spent only 6.3% of their time "taking decisions," while executives in centralized companies spent 14.6% of their time "taking decisions." He further found that executives in decentralized companies spent less time giving orders (6.8% versus 13.8%) than did executives in centralized companies. Carlson's finding that executives in decentralized companies gave fewer orders and made fewer decisions supports the claim that decentralization can lead to greater autonomy at the lower levels of the organization. However, there is a problem with this study that makes the results difficult to interpret. It is not clear what Carlson used as his criteria for classifying organizations as centralized or decentralized. Thus, it is impossible to know which particular type of decentralization Carlson's findings support. His findings are further limited by the fact that he had only nine managers as subjects in the study.

Weiss (1957) studied the relationship between centralized and decentralized organization structures and several measures of behavior. He classified 34 companies as either centralized or decentralized based upon their answers to a 22-item questionnaire about the level at which decisions were made in the company. Weiss found no significant differences between centralized and decentralized companies on any of the following variables: turnover rate, number of grievances, number of white collar workers, absenteeism, accident frequency, accident severity, and age of managers. It should be pointed out that although Weiss found no statistically significant differences, the trend on each of the factors considered was favorable to decentralized organizations.

Summary

The studies reviewed offer no clear support for the proposition that decentralization can produce either improved job attitudes or performance. Thus, Pfiffner and Sherwood's (1960) conclusion that decentralization must be accepted partially on faith seems completely justified. The fact that two of the studies reviewed obtained a few significant results and that another showed interesting trends suggests that the topic is a meaningful one for research. Perhaps the chief obstacle to research in this area, though, is the lack of an adequate method for measuring the degree of decentralization that exists in an organization. Until such a measure is developed the research evidence gathered will undoubtedly remain difficult to interpret.

It is doubtful that future research will ever prove that any of the three types of decentralization mentioned earlier are superior to centralization in all situations. Marschak (1959) has presented a model graphically illustrating that the efficiency of decentralization of decision making varies greatly depending upon the type of business in which the company is engaged. Sloan (1963) has made a similar point by stating that different degrees of decentralization appeared to be optimum for DuPont and General Motors because of their different product lines. Thus, for companies in a business where there is a high degree of complementarity among the members' actions (e.g., airlines or railroads), such that one member's action depends directly upon what his fellow employees are doing, a high degree of centralization is crucial. However, for companies in a business where there is a great deal of independence among the action of the members (e.g., chains of retail stores), such that one member's action does not depend on what his fellow employees are doing, increased decentralization may be desirable.

CONCLUSIONS

In this final section, two broad questions will be considered: First, what is the status of our current knowledge concerning the possible effects of structure? Second, where should future research efforts in this area be directed?

With respect to the first question, a review of the literature demonstrates the following:

1. Five of the seven properties of organization structure (span of control and centrali-

zation/decentralization being the two possible exceptions) have been shown to have some kind of significant relationship to either job attitudes or job behavior, or to both of these types of variables. However, as we stressed in the introduction, experimental "proof" of cause-effect relationships between structure and employee attitudes and behavior is elusive and almost nonexistent.

2. Certain structural variables seem to have stronger relationships to attitudes and behavior than other structural variables. On the basis of the evidence to date, the two properties of structure that have the strongest relationships with, or effects on, the two types of "dependent" variables are two suborganization properties: organizational levels and subunit size. Level was shown to have a definite connection with several types of attitudes, and subunit size was found to be associated strongly with both job satisfaction attitudes and several types of behavior. Three other properties—line/staff type of position, total-organization size, and tall/flat shape—accounted for some significant relationships to several of the dependent variables, but the strength and clarity of their relationships were not as great as was the case with organizational levels and subunit size. Each of these three former variables has not been researched as extensively as have the latter two variables, and hence it is possible that future results will show them to have somewhat stronger relationships than appears to be the case at the moment. The other two variables, span of control and centralization/decentralization, have so far not been found to be significantly related to attitudes or behavior. Again, however, relatively little research has been carried out on these two structural variables and they may yet be found to have more important influences on employees than seems apparent at the moment.

3. Certain dependent variables are more highly related to properties of structure than are other kinds of dependent variables. For example, indices of need satisfaction seem to be much more strongly correlated with structural properties than are indices of need importance. This makes good sense logically and psychologically, in that satisfactions presumably should be determined more by environmental factors (e.g., the degree of job responsibilties, supervisory behavior toward subordinates, size of work group, etc.) than by personal factors (e.g., degree of extroversion), whereas employee wants and desires are likely to come about more from personal factors than from aspects of the work environment. On the behavioral side, absenteeism and turnover seem somewhat more clearly related to structural factors than does employee output per se. The reasons why this should be so probably revolve around the fact that output or performance rates seem to be much more complexly determined phenomena than are absenteeism and turnover rates. That is, the latter types of behavior are within the employee's repertory of potential actions; however, all levels of output are not within his repertory of actions both because of personal limitations and because of possible environmental factors such as work-group pressure and equipment limitations. Therefore, if structure affects various aspects of behavior largely through its influence on employees' motivations, absenteeism and turnover rates will be more directly affected than will output levels.

In general, the impact of structural variables appears to be clearer on attitudes than on behavioral variables. However, this conclusion should be interpreted with caution because a search of the literature reveals that for most of the structural properties somewhat more research has been carried out on their relationships to attitudes than to types of behavior.

4. The direction of relationships of certain structural variables to certain dependent variables seems clear. The following directional relationships seem to be well supported by the research evidence to date: (*a*) organizational levels—positive relationship between height of level and degree of job and need satisfaction, positive relationship between height of level and perceived necessity for innerdirected type of job behavior; (*b*) line/staff hierarchies—positive relationship between line type of position and degree of need satisfaction; (*c*) subunit size—negative relationship between size and degree of job satisfaction, positive relationship between size

and absenteeism rates, positive relationship between size and turnover rates.

For certain other structural variables, the direction of their impact on certain dependent variables is not clear, even though particular studies found statistically significant relationships. Thus, it is not at all certain whether a flatter type of organization structure is always associated with greater employee satisfaction compared to a taller structure, or that a smaller total organization results in greater employee satisfaction than does a larger total organization. As will be pointed out below, the direction of relationship in each case seems to be affected by the interaction of some second structural variable. Also, the evidence on the relationship between subunit size and output shows a wide scattering of types of relationships, going all the way from negative through curvilinear to positive relationships. Thus, it may be that subunit size is an important determinant of employee productive behavior, but that the direction of its effects depends on relatively local and particularized conditions.

We now turn to our second question: Where should research efforts be directed in the future?

First, we would suggest that future research investigations in this area must be addressed to more complex questions. It seems evident that a great deal more attention has to be given to the possible interrelationships between and among different organization structural variables than has been the case so far. There is already some indication in the literature that such interactions among two or more properties or dimensions of structure are likely. For example, the available evidence indicates that in order to determine the effects of either subunit size or total-organization size, one must specify or take into account the organizational level or levels being considered. It seems probable that certain effects of size, either subunit or total, that are present when rank-and-file workers are the objects of investigation may be greatly reduced or possibly even reversed at management levels, especially middle and upper executive levels. With respect to a different structural variable, tall/flat organization shape, the available data indicate that its relationship to job satisfaction is modified considerably by the total size of the organizations being studied. In other cases there are, as yet, few data to support (or deny) a notion of interaction effects, yet logic would lead one to predict the likelihood of this kind of situation. As just one example, it appears probable that success, or lack of it, in cases of deliberate organizational decentralization may well depend upon which organizational levels are to be involved in the decentralization and upon such other intraorganizational structural variables as the spans of control and sizes of subunits created by the change.

To sum up this point, there are already enough indications in the literature to support a greater research effort to investigate the interactions among structural properties of organizations in their relationship to employees' job behavior and attitudes. Too much previous theorizing in the area of organizations has neglected such interaction possibilities, and hence there has been an unfortunate tendency to oversimplify vastly the effects of particular variables. Organizations appear to be much too complex for a given variable to have a consistent unidirectional effect across a wide variety of types of conditions.

Second, there is an obvious need for field experimental studies, as well as the more typical field comparative studies. We need, in other words, investigations where one or more structural variables are systematically manipulated in field, not laboratory, conditions. There are admittedly great practical difficulties in carrying out such investigations, yet with the study of organizations occupying such an important place in the behavioral sciences today, it is time that researchers in this area faced up to the necessities for this kind of undertaking. Perhaps this will require projects of greater scope and support than in the past, and most certainly it is likely to require a cross-discipline approach. But such projects are definitely not out of the realm of the possible. As examples, one can conceive without too much difficulty of the possibilities of manipulating such variables as span of control or size of subunit. The results of such studies would provide firm support for some aspects of organization theories, where such support is largely lacking at the present time.

Third, longitudinal studies are probably necessary to draw valid conclusions concerning certain structural variables. Results from studies so far carried out on the relationships of organizational level or line/staff hierarchies to job attitudes are subject to the limitation that any obtained relationships might in part be due to differences among types of people rather than to differences among types of positions within organizations. One profitable method of sorting out the relative effects of these two types of variables, that is, individual and organizational, would be to conduct longitudinal studies where the same individuals are followed in their careers from one kind of organizational position to another. Such longitudinal studies are conspicuous by their rarity in the literature on organization structure.

In conclusion, we feel that the importance attached to the study of organization structure is not misplaced. Increased attention in the future to research on structural properties of organizations should improve our understanding of the way people think and behave when they function in their jobs within organizations.

REFERENCES

Acton Society Trust. *Size and morale.* London: AST, 1953.

Argyle, M., Gardner, G., & Cioffi, I. Supervisory methods related to productivity, absenteeism, and labor turnover. *Human Relations,* 1958, **11**, 23–40.

Argyris, C. *Personality and organization.* New York: Harper, 1957.

Ash, P. The SRA Employee Inventory: A statistical analysis. *Personnel Psychology,* 1954, **7**, 337–364.

Baehr, Melany E. A factorial study of the SRA Employee Inventory. *Personnel Psychology,* 1954, **7**, 319–336.

Baker, Helen, & France, R. R. *Centralization and decentralization in industrial relations.* Princeton: Industrial Relations Section, 1954.

Baumgartel, H., & Sobol, R. Background and organizational factors in absenteeism. *Personnel Psychology,* 1959, **12**, 431–443.

Benge, E. J. How to learn what workers think of job and boss. *Factory Management and Maintenance,* 1944(May), **102**(5), 101–104.

Bennis, W. G. Leadership theory and administrative behavior: The problem of authority. *Administrative Science Quarterly,* 1959, **4**, 259–301.

Brayfield, A. H., & Crockett, W. H. Employee attitudes and employee performance. *Psychological Bulletin,* 1955, **52**, 396–424.

Brown, A. Some reflections on organization: Truths, half-truths, and delusions. *Personnel,* 1954, **31**(1), 31–42.

Browne, C. G., & Neitzel, Betty J. Communication, supervision, and morale. *Journal of Applied Psychology,* 1952, **36**, 86–91.

Burns, T. The directions of activity and communication in a departmental executive group: A quantitative study in a British engineering factory with a self-recording technique. *Human Relations,* 1954, **7**, 73–97.

Campbell, H. Group incentive payment schemes: The effects of lack of understanding and of group size. *Occupational Psychology,* 1952, **26**, 15–21.

Campbell, J. W. An attitude survey in a typical manufacturing firm. *Personnel Psychology,* 1948, **1**, 31–39.

Carlson, S. *Executive behavior, a study of the work load and the working methods of managing directors.* Stockholm: Stromberg, 1951.

Centers, R. Motivational aspects of occupational stratification. *Journal of Social Psychology,* 1948, **28**, 187–217.

Chandler, A. D. Management decentralization: An historical analysis. *Business History Review,* 1956, **30**, 111–174.

Cleland, S. *Influence of plant size on industrial relations.* Princeton: Princeton Univer. Press, 1955.

Coates, C. H., & Pellegrin, R. J. Executives and supervisors: Contrasting self-conceptions and conceptions of each other. *American Sociological Review,* 1957, **22**, 217–220.

Cordiner, R. J. *New frontiers for professional managers.* New York: McGraw-Hill, 1956.

Dale, E. *Planning and developing the company organization structure.* New York: American Management Association, 1952.

Dalton, M. Conflicts between staff and line managerial officers. *American Sociological Review,* 1950, **15**, 342–351.

Dalton, M. *Men who manage.* New York: Wiley, 1959.

Davis, K. Management communication and the grapevine. *Harvard Business Review,* 1953, **31**(5), 43–49.

Drucker, P. *Concept of the corporation.* New York: Day, 1946.

Dubin, R. Business behavior *behaviorally* viewed. In G. B. Strother (Ed.), *Social science approaches to business behavior.* Homewood, Ill.: Dorsey Press, 1962.

Dubin, R., & Spray, S. L. Executive behavior and interaction. *Industrial Relations,* 1964, **3**, 99–108.

Entwisle, Doris, & Walton, J. Observations on the span of control. *Administrative Science Quarterly,* 1961, **5**, 522–533.

Fisch, G. G. Line-staff is obsolete. *Harvard Business Review,* 1961, **39**(5), 67–79.

Fisch, G. G. Stretching the span of management. *Harvard Business Review,* 1963, **41**(5), 74–85.

Fleishman, E. A. The measurement of leadership attitudes in industry. *Journal of Applied Psychology,* 1953, **37**, 153–158.

Fleishman, E. A., Harris, E. F., & Burtt, H. E. Leadership and supervision in industry: An evaluation of a supervisory training program. *Ohio State University, Bureau of Educational Research Monograph,* 1955, No. 33.

Fortune. The Fortune survey: A self-portrait of the American people—1947. *Fortune,* 1947, **35**(1), 5–16.

Fortune. Effective morale. *Fortune,* 1950, **42**(2), 46–50.

Gardner, B. B., & Moore, D. G. *Human relations in industry.* (3rd ed.) Homewood, Ill.: Dorsey Press, 1955.

Ginzberg, E., & Reilley, E. W. *Effecting change in large organizations.* New York: Columbia Univer. Press, 1957.

Given, W. B. *Bottom-up management; people working together.* New York: Harper, 1949.

Gordon, G., & Becker, S. Organizational size and managerial succession: A reexamination. *American Journal of Sociology,* 1964, **70**, 215–223.

Graicunas, V. A. Relationship and organization. In L. Gulick & L. Urwick (Eds.), *Papers on the science of administration.* New York: Institute of Public Administration, 1937.

Grusky, O. Corporate size, bureaucratization, and managerial succession. *American Journal of Sociology,* 1961, **67**, 261–269.

Haire, M. *Psychology in management.* New York: McGraw-Hill, 1956.

Haire, M., Ghiselli, E. E., & Porter, L. W. Cultural patterns in the role of the manager. *Industrial Relations,* 1963, **2**, 95–117.

Handyside, J. D. Satisfactions and aspirations. *Occupational Psychology,* 1961, **35**, 213–244.

Hemphill, J. K. Relations between the size of the group and the behavior of "superior" leaders. *Journal of Social Psychology,* 1950, **32**, 11–22.

Hemphill, J. K. Group dimensions: A manual for their measurement. *Ohio State University, Bureau of Business Research Monograph,* 1956, No. 87.

Herbst, P. G. Measurement of behavior structure by means of input-output data. *Human Relations,* 1957, **10**, 335–346.

Herzberg, F., Mausner, B., Peterson, R. O., & Capwell, Dora F. *Job attitudes: Review of research and opinion.* Pittsburgh: Psychological Service of Pittsburgh, 1957.

Herzberg, F., Mausner, B., & Snyderman, Barbara B. *The motivation to work.* (2nd ed.) New York: Wiley, 1959.

Hewitt, D., & Parfitt, Jessie. A note on working morale and size of group. *Occupational Psychology,* 1953, **27**, 38–42.

Hull, R. L., & Kolstad, A. Morale on the job. In G. B. Watson (Ed.), *Civilian morale.* New York: Houghton Mifflin, 1942.

Indik, B. P., & Seashore, S. E. *Effects of organization size on member attitudes and behavior.* Ann Arbor: University of Michigan, Survey Research Center of the Institute of Social Research, 1961.

Kahn, R. L. Human relations on the shop floor. In E. M. Hugh-Jones (Ed.), *Human relations and modern management.* Amsterdam: North-Holland Publishing, 1958. Pp. 43–74.

Katzell, R. A., Barrett, R. S., & Parker, T. C. Job satisfaction, job performance, and situational characteristics. *Journal of Applied Psychology,* 1961, **45**, 65–72.

Kerr, W. A., Koppelmeier, G. J., & Sullivan, J. J. Absenteeism, turnover and morale in a metals fabrication factory. *Occupational Psychology,* 1951, **25**, 50–55.

Kolstad, A. Attitudes of employees and their supervisors. *Personnel,* 1944, **20**, 241–250.

Kreisberg, L. Careers, organization size, and succession. *American Journal of Sociology,* 1962, **68**, 355–359.

Lawler, E. E., & Porter, L. W. Perceptions regarding management compensation. *Industrial Relations,* 1963, **3**, 41–49.

Lawrence, P. R. *The changing of organizational behavior patterns: A case study of decentralization.* Boston: Harvard University, Graduate School of Business Administration, Division of Research, 1958.

Leavitt, H. J. *Managerial psychology.* Chicago: Univer. Chicago Press, 1958.

Leavitt, H. J., & Whisler, T. L. Management in the 1980's. *Harvard Business Review,* 1958, **36**(6), 41–48.

Likert, R. *New patterns of management.* New York: McGraw-Hill, 1961.

Litzinger, W. D. Entrepreneurial prototype in bank management. *Academy of Management Journal,* 1963, **6**(1), 36–45.

McGregor, D. *The human side of enterprise.* New York: McGraw-Hill, 1960.

Mandell, M. M. *Recruiting and selecting office employees.* New York: American Management Association, 1956.

Marriott, R. Size of working group and output. *Occupational Psychology,* 1949, **23**, 47–57.

Marschak, J. Efficient and viable organizational forms. In M. Haire (Ed.), *Modern organization theory.* New York: Wiley, 1959. Pp. 307–320.

Martin, N. H. The levels of management and their mental demands. In W. L. Warner & N. H. Martin (Eds.), *Industrial man.* New York: Harper, 1959. Pp. 276–294.

Meltzer, L., & Salter, J. Organizational structure and the performance and job satisfaction of physiologists. *American Sociological Review,* 1962, **27**, 351–362.

Metzner, Helen, & Mann, F. Employee attitudes and absences. *Personnel Psychology,* 1953, **6**, 467–485.

Morse, Nancy C. *Satisfactions in the white-collar job.* Ann Arbor: University of Michigan, 1953.

Opinion Research Corporation. *The conflict between the scientific and the management mind.* Princeton: ORC, 1959.

OPINION RESEARCH CORPORATION. *Motivating managers.* Princeton: ORC, 1962.

PELLEGRIN, R. J., & COATES, C. H. Executives and supervisors: Contrasting definitions of career success. *Administrative Science Quarterly,* 1957, 1, 506–517.

PFIFFNER, J. M., & SHERWOOD, F. P. *Administrative organization.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

PORTER, L. W. A study of perceived job satisfactions in bottom and middle management jobs. *Journal of Applied Psychology,* 1961, **45**, 1–10.

PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology,* 1962, **46**, 375–384.

PORTER, L. W. Job attitudes in management: II. Perceived importance of needs as a function of job level. *Journal of Applied Psychology,* 1963, **47**, 141–148. (a)

PORTER, L. W. Job attitudes in management: III. Perceived deficiencies in need fulfillment as a function of line vs. staff type of job. *Journal of Applied Psychology,* 1963, **47**, 267–275. (b)

PORTER, L. W. Job attitudes in management: IV. Perceived deficiencies in need fulfillment as a function of size of company. *Journal of Applied Psychology,* 1963, **47**, 386–397. (c)

PORTER, L. W. Where is the organization man? *Harvard Business Review,* 1963, **41**(6), 53–61. (d)

PORTER, L. W. *Organizational patterns of managerial job attitudes.* New York: American Foundation for Management Research, 1964.

PORTER, L. W., & HENRY, MILDRED M. Job attitudes in management: V. Perceptions of the importance of certain personality traits as a function of job level. *Journal of Applied Psychology,* 1964, **48**, 31–36. (a)

PORTER, L. W., & HENRY, MILDRED M. Job attitudes in management: VI. Perceptions of the importance of certain personality traits as a function of line vs. staff type of job. *Journal of Applied Psychology,* 1964, **48**, 305–309. (b)

PORTER, L. W., & LAWLER, E. E. The effects of tall vs. flat organization structures on managerial job satisfaction. *Personnel Psychology,* 1964, **17**, 135–148.

PORTER, L. W., & SIEGEL, J. The effects of tall vs. flat organization structures on managerial satisfactions in foreign countries. Unpublished manuscript, University of California, Berkeley, 1964.

REVANS, R. W. Human relations, management and size. In E. M. Hugh-Jones (Ed.), *Human relations and modern management.* Amsterdam: North-Holland Publishing, 1958. Pp. 177–220.

ROSEN, H. Desirable attributes of work: Four levels of management describe their job environments. *Journal of Applied Psychology,* 1961, **45**, 156–160. (a)

ROSEN, H. Managerial role interaction: A study of three managerial levels. *Journal of Applied Psychology,* 1961, **45**, 30–34. (b)

ROSEN, H., & WEAVER, C. G. Motivation in management: A study of four managerial levels. *Journal of Applied Psychology,* 1960, **44**, 386–392.

SEASHORE, S. E. *Group cohesiveness in the industrial work group.* Ann Arbor: University of Michigan, Institute for Social Research, 1954.

SELZNICK, P. *T.V.A. and the grass-roots: A study in the sociology of formal organization.* Berkeley: Univer. California Press, 1949.

SHARTLE, C. L. *Executive performance and leadership.* Englewood Cliffs, N. J.: Prentice-Hall, 1956.

SLOAN, A. P. My years with General Motors—Part II. *Fortune,* 1963, **68**(4), 145–148.

SMITH, G. A., JR. *Managing geographically decentralized companies.* Boston: Harvard University, Graduate School of Business Administration, Division of Research, 1958.

STRAUSS, G., & SAYLES, L. R. *Personnel: The human problems of management.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

TALACCHI, S. Organization size, individual attitudes and behavior: An empirical study. *Administrative Science Quarterly,* 1960, **5**, 398–420.

TRIANDIS, H. C. Categories of thought of managers, clerks, and workers about jobs and people in an industry. *Journal of Applied Psychology,* 1959, **43**, 338–344. (a)

TRIANDIS, H. C. Differential perception of certain jobs and people by managers, clerks, and workers in industry. *Journal of Applied Psychology,* 1959, **43**, 221–225. (b)

TRIANDIS, H. C. A comparative factorial analysis of job semantic structures of managers and workers. *Journal of Applied Psychology,* 1960, **44**, 297–302.

URWICK, L. F. Executive decentralization with functional coordination. *Management Review,* 1935, **24**, 355–368.

VITELES, M. S. *Motivation and morale in industry.* New York: Norton, 1953.

WEISS, E. C. Relation of personnel statistics to organization structure. *Personnel Psychology,* 1957, **10**, 27–42.

WHYTE, W. F. *Men at work.* Homewood, Ill.: Dorsey Press, 1961.

WHYTE, W. H., JR. *The organization man.* New York: Simon & Schuster, 1956.

WOODWARD, JOAN. *Management and technology.* London: Her Majesty's Stationery Office, 1958.

WORTHY, J. C. Organizational structure and employe morale. *American Sociological Review,* 1950, **15**, 169–179.

WORTHY, J. C. *Big business and free men.* New York: Harper, 1959.

(Received September 22, 1964)

Psychological Bulletin
1965, Vol. 64, No. 1, 52–64

EXPERIMENTAL ATTEMPTS TO PRODUCE PHENYLKETONURIA IN ANIMALS:

A CRITICAL REVIEW[1]

RATHE KARRER AND GLENN CAHILLY[2]
Research Department, Training School at Vineland, New Jersey

Studies attempting to produce PKU in animals are extensively reviewed. While some studies claim to have produced and consequently even prevented PKU in animals, such claims were found to be unjustified at the present stage of research. Behavioral and biochemical criteria for assessing PKU are discussed along with certain methodological problems inherent in such work. The experimental production of PKU, on the other hand, offers an important model for the study of the relations between development, intellectual functioning, and biochemistry.

Phenylketonuria (PKU) is considered an "inborn error of metabolism" produced by an autosomal recessive gene. The primary defect is known to be a deficiency of the liver enzyme phenylalanine hydroxylase which results in reduced conversion of phenylalanine to tyrosine with a consequent buildup of phenylalanine and its degradation products. The defect does not become apparent until shortly after birth when production of the enzyme normally begins. At this time the brain is undergoing its most rapid rate of biochemical differentiation. This biochemical syndrome usually results in severe mental deficiency by as yet unknown processes. There is evidence that the defect can be controlled and the resultant mental deficiency averted if the patient is placed on a diet low in phenylalanine within the first months of life. As Knox (1963) points out, the situation is one of abnormal growth and development in an abnormal chemical milieu rather than one of toxic effects of abnormal metabolites. Thus, the question concerns the nature of the alterations induced by the abnormal chemical milieu in the developing brain rather than the effect of abnormal metabolites on the normal, developed brain.

Although PKU occurs relatively rarely (approximately 1:10–25,000) and accounts for less than 1% of mental retardates, it has attracted much attention and research of late. Its importance is primarily as a model for the study of the relations between genetics, biochemistry, and behavior. Within the past few years there have been a number of attempts to experimentally produce PKU in animals (cf. Waisman, 1963) and thus bring these relations under experimental manipulation. The realization of this goal is of some importance not only to those concerned with mental retardation but also to those concerned with the interrelations of normal cognitive functions, biochemistry, and development. Since this work is perhaps less known outside of mental retardation research, the present paper will review in some detail and try to evaluate critically these experimental studies. Accordingly, with few exceptions, only studies whose main purpose was to attempt to produce PKU judged by biochemical criteria and/or behavioral criteria will be included.

Attempts to produce PKU in animals have utilized three methods: (*a*) the administration of large doses of phenylalanine alone or in combination with other amino acids, (*b*) the administration of presumed phenylalanine hydroxylase inhibitors, and (*c*) by selective breeding. All of the methods attempted to create a condition of excess plasma phenylalanine, excretion of phenylpyruvic acid, and retardation, the most striking features of the human condition. It should be noted, however, that these specific biochemical features may

---

[1] The authors would like to thank George Rendina for initiating the interest in this area and for many informative discussions of the problems. Thanks are also expressed to Johs. Clausen for his cooperation and pertinent comments in the organization of the problem, and to Ethel Weiss for comments on the manuscript.

[2] Now at Worthington Biochemical Corporation, Freehold, New Jersey.

or may not be directly involved in the resultant brain dysfunction. Of further significance is the fact that these studies begin with animals having normal phenylalanine hydroxylase activity in contrast to the human condition.

STUDIES ON RODENTS

Waisman and his colleagues at Wisconsin, as well as Wassmuth in Germany, were the first to pose the question of experimentally producing PKU and thereby making possible experimental study.

Auerbach, Waisman, and Wyckoff (1958) reasoned that perhaps administration of the product (*L*-tyrosine) of the hydroxylation reaction of phenylalanine could result in an accumulation of substrate and elevated phenylalanine levels in the organism. They began a series of experimental studies on this basis by adding large amounts (5%) of *L*-tyrosine, and *DL*-phenylalanine alone or together to the diet of lactating or pregnant rats. While initial mortality of the weanling, newborn, or suckling animals was a problem, either amino acid alone, or together in reduced quantities (2.5 or 3.75%), allowed normal growth. The results were inhibition of phenylalanine hydroxylase activity to about one-third normal activity as well as more phenylpyruvate in the urine which reached maximal levels 1 month after start of the diet. There was, however, little effect on plasma phenylalanine levels. This was interpreted as substantiating their hypothesis that *L*-tyrosine, as well as *DL*-phenylalanine, inhibited phenylalanine hydroxylase activity.

By operant conditioning of licking a dry dipper reinforced by water on a 2-minute fixed-interval schedule, they attempted to assess whether their animals had any behavioral impairment. Using animals on 2.5% *L*-tyrosine and DL-phenylalanine diets and controls, with water deprivation as the motivator, they determined from the cumulative records the extent to which each group increased their licking in 2-minute cycles—"temporal discrimination." They found the control group increased their licking for 82.9% of the 2-minute intervals, while the diet groups increased only 46.8% (significant by *t* at .01). Since there was no pair-fed control group, it is impossible to determine whether this behavioral deficit was a direct consequence of the PKU diet or a result of simple malnutritional weakness. Although phenylalanine hydroxylase activity of tested animals was diminished to one-third normal, the plasma phenylalanine was raised only slightly (1.6 mg/100 ml to 2.8 mg/100 ml). Thus the biochemical criteria for PKU were not adequate. It should be noted that recent data (Waisman, 1960) indicate that *L*-tyrosine is unable to raise phenylalanine levels by inhibition of phenylalanine hydroxylase and, in fact, increases the enzyme's activity.

Neill and Langford (1961) reported no success in simulating PKU biochemically in rats by 18 days on a 2.5% *L*-phenylalanine or a 2.5% *L*-phenylalanine plus 2.5% *L*-tyrosine diet. Similar results were obtained for dosages of 3.75% *L*-phenylalanine plus 3.75% *L*-tyrosine for 33 days. Neither were they able to simulate PKU biochemically with a diet of 1% beta-2-thienylalanine (a phenylanine hydroxylase inhibitor) plus 1% *L*-phenylalanine for 3 weeks. These results are directly counter to those reported by Auerbach et al. (1958).

Yuwiler and Louttit (1961) and Louttit (1962), investigating the effects of a phenylalanine diet on brain serotonin and learning, used weanling rats divided by split litters into control and experimental groups. Dosages of phenylalanine administered in the diet ranged from 5–10%. The groups were run on an open field test at 46 days of age and placed on a 23-hour food deprivation at 49 days of age. They were then tested on the Hebb-Williams series of mazes and on a black-white successive discrimination with the basic control diet plus sugar as reward. After 21 days the animals were decapitated and the brains analyzed for serotonin. It was found that the phenylalanine animals had significantly lower serotonin (.468 μgm/gm tissue vs. .584 μgm/gm) than controls. It was also reported that the behavioral tests were "in the direction of" greater errors on the Hebb-Williams mazes (Louttit reports this as significant using an increased *N*) and *fewer* trials to criterion on the successive discrimination problem for the phenylalanine animals. There were no differences on the open field test. Louttit concludes that the results do not demonstrate "mental deficiency because of the better discrimination performance and the possibility that poor maze performance may also be due to increased exploratory behavior." One must agree with him, since in addition to the possible motivational differences there were no biochemical analyses of serum phenylalanine or urine phenylpyruvate. Thus none of the indicators of PKU were obtained.

Huang, Tannenbaum, Blume, and Hsia (1961) studied the metabolism of 5-hydroxyindole compounds in weanling rats fed the following three combinations of phenylalanine and/or tyrosine: (*a*) 3.75% *L*-tyrosine plus 3.75% *DL*-phenylalanine for 2 and 3 months; (*b*) 3.75% *L*-tyrosine and 2.5% *DL*-phenylalanine for 2 months; (*c*) 2.5% *DL*-phenylalanine for 2 months. Urine samples were taken weekly and blood samples by cardiac puncture at 6.5, 7, and 9 weeks after start of the diet. The urine was analyzed for phenylpyruvic acid and 5-hydroxyindolacetic acid, blood plasma for phenylalanine and 5-hydroxytryptamine (serotonin), and the liver for phenylalanine hydroxylase activity. Blood plasma phenylalanine and urine phenylpyruvic acid of the experimental animals were higher than controls for diet No. 1 and markedly higher for No. 2, while there was little difference for No. 3. Similarly, 5-hydroxyindolacetic acid and 5-hydroxytryptamine were decreased. Return to a normal diet reversed these effects. They concluded that PKU can be induced by *DL*-phenylalanine plus *L*-tyrosine, while *DL*-phenylalanine alone gives only a slight effect. It is difficult to accept this conclusion since no pair-fed controls were used, the levels of plasma phenylalanine were low (equivalent to 4.24 mg.% on diet No. 2), there were no behavioral indices, and the effects observed were reversible.

Loo, Diller, and Owen (1962) fed pregnant rats 1% *DL*-phenylalanine and then placed the pups on the same diet at 4 weeks of age. At 16 weeks of age they were given seven sessions of 160 trials each on a conditioned avoidance task. Blood from the tail was analyzed for amount of keto acids and for phenylalanine. No difference in keto acids were found between experimental and control groups, but phenylalanine level was reported to be significantly different (61 ± 5 and 50 ± 3 μgm/ml for phenylalanine and control groups, respectively, $p = 0.1$). This is equivalent to 6.1 and 5.0 mg.%, hardly comparable to levels found in human PKU. The phenylalanine group also failed to learn the conditioned avoidance task. Forty-five days after the first test they were given a second learning session after having been on a normal diet for 20 days. Four of the six phenylalanine animals then succeeded in learning the task, and their blood phenylalanine levels were normal. With such small amounts of phenylalanine and no pair-fed controls, it is difficult to accept these results as directly due to phenylalanine. The nonpermanence of the effects when placed on a normal diet at adulthood also argues against a PKU defect.

McKean, Schanberg, and Giarman (1962) and Schanberg, McKean, and Giarman (1962) reported unpublished data that rats "made phenylketonuric" by a diet containing 3.5% phenylalanine and 3.5% tyrosine for 14 days had a greater number of errors and slower performance in swimming a water maze and "high circulating levels of phenylalanine and tyrosine." They also reported the reversal of this effect when the phenylketogenic diet was supplemented with 5% tryptophan or 7% glucose. No further information was given with which to assess the validity of the behavioral measures or to determine the adequacy of the biochemical criteria.

Wang and Waisman (1961) obtained additional information on effects of feeding *L*-phenylalanine, since the *D*-isomer previously used was recognized as being deaminated by *D*-amino acid oxidase in the kidney to give large amounts of phenylpyruvic acid in the urine without the metabolic byproducts found in PKU. They fed 0, 3.5, 5, or 7% *L*-phenylalanine in a ground commercial chow to 18-day-old weanling rats for 3 weeks. In an attempt to manipulate feeding habits so as to maintain more uniform phenylalanine serum levels, some rats were kept under cycling 3-hour light and dark conditions. The rats on a 5 or 7% *L*-phenylalanine diet showed little effect of the light conditions, apparently eating half that of normals at more frequent intervals regardless of lighting. This reduction of intake was considered as probably due to *L*-phenylalanine. Biochemically they found that 3.5% neither affected growth nor appreciably influenced plasma phenylalanine levels or gave increased phenylketone excretion. Five percent seriously affected growth and resulted in elevated plasma phenylalanine, especially in light-dark housed animals, but gave inconsistently increased excretion of urinary phenylketones. Seven percent *L*-phenylalanine even more adversely affected growth (10% mortality in some groups) and resulted in plasma phenylalanine levels comparable to PKU humans (range from 12–70 mg.%) as well as consistently raised urinary phenylketones. It was also found that regardless of phenylalanine dosage, the phenylalanine hydroxylase activity of the liver was reduced 40–60%. It is indicated that *L*-phenylalanine dosage is more important in attaining high serum levels than duration of the experimental or light conditions. While no behavioral data are given, such work is reportedly in progress.

Boggs, Polidora, and Waisman (1962) reported that pregnant and lactating rats fed diets containing 5 and 7% *L*-phenylalanine were apparently able to adapt to high phenylalanine in that almost normal plasma levels were regained after several weeks. They noted that while the free phenylalanine in the milk was 20–40 times higher than normal, there was no significant effect on plasma phenylalanine levels or liver phenylalanine hydroxylase activity. This was attributed to a relatively high protein level of rat milk which makes a large increase in phenylalanine insignificant in terms of total phenylalanine. Boggs and Waisman (1962) report data indicating increased plasma levels in suckling pups from females fed 7% *L*-phenylalanine or 4% *DL* plus 3% *L*-phenylalanine. Plasma tyrosine was also found to be low in the 7% *L*-phenylalanine group. These plasma levels were not comparable to human PKU and no phenylketones were detected in the urine. Boggs and Waisman (1964a) reported the elevation of plasma phenylalanine and lowered phenylalanine hydroxylase activity in pregnant female albino rats fed 5 and 7% *L*-phenylalanine. There was, however, large individual variation. Similar determinations in fetal rats from these pregnant females indicated plasma phenylalanine levels comparable to human PKU, but again with great individual variation (mean plasma phenylalanine and standard deviations were 4.9 ± 1.0 and 30.5 ± 28.3 for control and 7% phenylalanine groups, respectively). The fetal plasma phenylalanine levels remained about twice that of the dam regardless of the latter's level. They reported no gross abnormalities in such pups regardless of the increased fetal plasma phenylalanine, but feel that it is still likely that *in utero* brain damage did occur. Boggs and Waisman (1964b) reported higher plasma phenylalanine and tyrosine, lower brain serotonin, and higher urinary phenylketones in rats fed 5 or 7% *L*-phenylalanine from weaning to about 50–60 days of age compared to pair-fed or ad lib controls. Phenylalanine-pyruvate transaminase activity was apparently increased significantly in the pair-fed groups ($t = 3.73$, $p < .001$, calculated by the writer), but not in the phenylalanine group compared to the ad lib group. Phenylalanine hydroxylase activity was again shown to be reduced about 50% when compared to ad lib controls, but no data is given to allow comparison to pair-fed controls. There was an adaptation-like effect in that there was little difference between the groups on these measures at 80–90 days. This series of studies has had limited success in producing the biochemical correlates of PKU in rats and no behavioral data are provided.

Polidora, Boggs, and Waisman (1963; also Boggs & Polidora, 1963), in the most complete report of behavioral assessment to date from the Wisconsin group, described a behavioral testing procedure which they claim to be reliable in detecting impaired maze performance of PKU rats. Since they recognized that a task using a food- or water-motivating system may confound the situation, they chose escape from water as a motivator in a water maze. They used a 6-unit, water-filled T maze (cf. Schanberg, McKean, & Giarman, 1961) and 21-day-old rats on a 5% *L*-phenylalanine diet for the first 10 days followed by 7% (cf. Wang & Waisman, 1961) until the end of the experiment. Control rats received the same commercial diet without *L*-phenylalanine. Behavioral testing was done on three different groups aged 72, 87, and 90 days. After training in a 50-inch straight channel (5 trials) on Day 1 (pretest), the rats were given 7 trials per day (Days 2, 3, 4) in the maze followed by another 5 trials in the straight channel (Day 5, posttest) comparable in length to the maze (100 inches). Successive trials were spaced 10 minutes apart. Time between start and finish was recorded as well as "transit speed" (linear extent/elapsed time) and errors (whole body cul entry). Transit speeds were found to be slower on the posttest for the experimental groups. The authors state that this is due to more backtracking and circling by the experimental rats which, when taken into consideration, gave swimming speeds "practically identical for the two groups." Analysis of variance of transit speeds as well as errors on the maze trials (Days 2, 3, 4) indicated a significant difference between experimental and control groups. This was complicated by significant effects of trials, days, and trials-by-days interaction. In essence, the experimental groups performed more slowly and made more errors. The slower performance resulted mainly from their higher error rate, slowness in gaining exit from a cul, and a tendency to turn in the wrong direction upon exit. It is also reported that the experimental rats had 38% lower body weight and infrequent mild alopecia, although there was no apparent muscular weakness. While a clear-cut behavioral deficit seems to have been demonstrated, it is not apparent whether it is the result of excess *L*-phenylalanine per se rather than from a self-imposed starvation or malnutrition. Such a demonstration requires more adequate controls such as pair-feeding and "innocuous" control diets, resulting in similar self-imposed starvation.

In this respect it is of interest that Boggs, Rosenberg, and Waisman (1963) have reported that feeding 5 or 7% *L*-phenylalanine caused a decrease in brain serotonin even compared to pair-fed controls. A decrease in liver serotonin was found in both groups and attributed to the restricted food intake. The brains of the pair-fed controls, however, were able to maintain normal serotonin levels while the liver levels decreased, a result which is not understood.

Perez and Schmidt (1963), in a pilot study using albino rats and intraperitoneal injection, administered 1.0, 1.5, and 2.0 gm. of *L*-phenylalanine per kg. of body weight dissolved in saline (2.5 mg/ml) to neonate rats from about 2 hours after birth for 60 consecutive days. Control rats received comparable volumes of saline. Ferric chloride urine tests were positive for the 1.5 and 2.0 mg/kg groups as early as 3 hours after birth. A significant cumulative depression of growth occurred during the phenylalanine administration but was reversible afterwards. Of great importance is the mortality rate which ranged from 36% for the controls to 70% for the highest dosage group. Behavioral measures were obtained 46–64 days after the last injection and after a 10-day adaptation to a 23.5-hour water-deprivation schedule. Animals were trained in a straight runway, 10 runs a day for 5 days. Each animal was then given 10 trials a day in a Lashley III maze until reaching a criterion of 9/10 errorless trials. They found that male experimental animals made significantly more errors and tended to require more trials. While this method has been found to raise serum phenylalanine for about 10 hours after injection (peak of about 160 mg/% after 2 hours[3]), more extensive recent attempts to replicate these behavioral results have been negative (Perez, 1964).[4] It is suggested that behavioral impairment found during phenylalanine administration may be due to toxic effects of phenylalanine rather than true PKU. One weakness of these results is the possible selective factor of mortality; only the biggest and best survive. To be emphasized, however, is that the behavioral testing was done after ample time to recover from the specific administration effects with the pertinent rational that to adequately stimulate PKU the effect must be permanent. This allows the use of a motivator compatible with standard behavioral assessment techniques without confounding the situation, but does not obviate controls (difficult in this case) to determine the effects of the self-imposed diet restrictions of the amino acid imbalance animals.

---

[3] V. J. Perez. Phenylpyruvic oligophrenia in the rat. II. Quantification of serum *L*-phenylalanine. Personal communication, January 1965.

[4] V. J. Perez. Phenylpyruvic oligophrenia in the rat. I. Personal communication, January 1965.

Woolley and Hoeven (1963) briefly reported that weanling mice fed 3.5% *DL*-phenylalanine plus 3.5% *L*-tyrosine gave similar changes in urinary excretion of phenylpyruvic acid as shown by Auerbach et al. (1958) and in serotonin and catechol amines of the tissues as shown by Huang et al. (1961) and Wang, Harwalkar, and Waisman (1962). They also reported a slight *increase* in learning ability measured by a T maze with escape from light as the motivator. Needless to say, this is counter to the results of Auerbach et al. (1958) for rats performing a different task, but similar to that reported by Louttit (1962) for discrimination learning. They explain this result by pointing out that the crucial period in phenylketonuria is early infancy, whereas the animals used in the study were "adults." They claim evidence (Woolley, 1962) that it is a serotonin deficiency induced early in life which damages the intellect, while deficiency induced in adulthood somehow enhances learning ability. It is of interest that they report that the observed behavioral changes were not permanent, the mice obtaining normal scores several days after cessation of treatment. In another study (Woolley & Hoeven, 1964a) they began force-feeding mice within 24 hours of birth .01 ml. of *DL*-phenylalanine (20 mg/ml) and *L*-tyrosine (10 mg/ml) nine times per day. The dosage was increased with growth so that at 2 weeks of age they were receiving .1 ml/hr, at which time force-feeding was stopped and 3.5% each of *DL*-phenylalanine and *L*-tyrosine added to the food. At 7–8 weeks of age they were placed on a normal diet for 3 days and then tested in a T maze as in the previous study. Compared to two groups of control animals, one force-fed water followed by a normal diet and the other reared normally, the phenylalanine-treated animals did more poorly and exhibited phenylpyruvic acid in the urine. Other animals given reserpine or chlorpromazine to produce serotonin deficiency during development performed similarly. They report that the phenylalanine group was physically normal at maturity but tended to be irritable and less active. Further studies (Woolley & Hoeven, 1964b, 1964c) reported that when serotonin cogeners (melatonin and HTP) were administered concurrently with the phenylalanine diet from birth, "the mental defect was prevented even though excesses of phenylalanine and . . . phenylpyruvic acid still remained in the tissues."

This is an impressive series of studies, having not only claimed the production of the mental defect of phenylketonuria in mice but also its prevention. The claim of mental defect is based upon trials to learn a T maze, which data are amazingly tight and consistent as reported but which are somewhat misleading for assessing the significance between experimental treatments (not reported). The standard deviations reported reflect the variance of a population of subgroup means rather than the variance within each experimental treatment. This latter variance among individual mice is considerable, which would lead one to suspect the significance of the small mean differences obtained (generally about one trial in a test of maze learning, e.g., 7.4 correct out of 10 trials for normals vs. 8.3 correct out of 10 trials for phenylalanine mice). The fact that these authors are apparently able to consistently reproduce these small differences, however, indicates some real difference between treatments. McMillan (1964) has recently emphasized that the behavioral effects were not permanent but restricted to the period of treatment, suggesting impaired performance rather than learning. That no analyses other than the presence of phenylpyruvate in the urine were done in order to assess better the biochemical simulation of PKU also weakens these studies, since the diet used has not been successful with rats (see above). The phenylpyruvate may very well be due to the nonphysiological *D*-isomer which is converted in the kidney. Other factors such as the lack of pair-feeding might also play a role. On the other hand, the T-maze task using light avoidance and the later use of shock-avoidance learning seem well suited to this problem, especially when time on a normal diet is allowed before testing. These experiments need to be closely replicated by other investigators before they can be fully evaluated.

Wassmuth (1958) and Ammon (1961) report an attempt to produce PKU in rabbits by daily subcutaneous (261 mg.) or intravenous

(131 mg.) injection of *DL*-phenylalanine before copulation, during pregnancy, and during lactation. Of 14 offspring, 2 died at birth, 3 had partial paralysis and subsequently died, while the rest were described as "unruly" and "nervous" with unhealthy, "shaggy and patchy" fur, especially on the head. Even though their EEGs were normal and no significant increase of phenylpyruvic acid in the urine was found, the author concludes that these young exhibited "disorders comparable with those of human phenylketonuria." With the use of the nonphysiological isomer, no controls, and scanty data it is difficult to accept such a conclusion.

An interesting study by Coleman (1960) suggests another approach to experimental PKU. It was reported that certain strains of mice (DBA/1J and DBA/2J) were characterized by dilute pigmentation, extreme nervousness, and seizures. These mice had a diminished phenylalanine hydroxylase activity ranging from 14–86% of that of an unrelated strain (C57BR/cdj). Evidence was found for the presence of an enzyme inhibitor, possibly phenylacetic acid. Of further interest is the recent report by Wimer and Weller (1964) that the DBA/2J strains did poorer on a black-white discrimination task than a C57BL/6J strain. These studies are suggestive enough to make further research with this approach appear promising.

Watt and Vandevoorde (1964) have recently reported that rats fed .2% *DL*-4-fluorophenylalanine, a competitive inhibitor of phenylalanine hydroxylase, excreted approximately five times more phenylpyruvic acid than controls. The use of enzyme inhibitors alone or together with excess phenylalanine may be a productive approach, although there are apparently no known completely specific inhibitors of phenylalanine hydroxylase.

STUDIES ON MONKEYS

Waisman, Wang, Palmer, and Harlow (1960, 1962; Waisman, 1962) report a lengthy, ongoing study of four macaque monkeys separated from their mothers as soon as possible and reared on surrogate mothers (cf. Harlow, 1958). They were given increasing dosages (up to .5 gm/100 ml) of *L*-phenylalanine from soon after birth for 15 months. The average daily intake of *L*-phenylalanine approximated 2.75 gm/kg body weight. Regularly determined blood phenylalanine indicated generally high plasma phenylalanine, but the consistency of elevation was quite variable in each animal (interpolations from their figures give ranges from approximately 1 mg.% to 50 mg.%). The authors suggest that this frequent fall in level is probably due to poor food intake. Throughout the study the ferric chloride test, 2, 4-dinitrophenylhydrazine test, or the Phenistix test as well as quantitative analysis, strongly indicated the presence of phenylketones in the urine. Behavioral evaluation of delayed reaction, object discrimination, the Hebb-Williams maze series, and discrimination-learning set (cf. Harlow, 1959) indicated poorer performance on delayed response and possibly the Hebb-Williams mazes and discrimination-learning set. Individual scores for the four experimental animals are contrasted, apparently, to a mean score of an indeterminant number of normal controls. It would be important to know if the differences observed were significant in view of the variation among the four experimental animals (there is no mention of normal variation). Nor is there mention of how the normal controls (presumably reared under identical conditions of isolation with surrogate mothers) would perform compared to normally reared monkeys. In other words, what is the role of the rearing conditions in the behavioral deficit on these tasks, especially since such conditions are known to produce deficits in other behavioral dimensions (Harlow, 1962)? Green and Gordon (1964), for example, have recently demonstrated that monkeys reared in isolation show less interest in looking at such objects as food, geometric forms, the mothers, peers, or an empty chamber. Although no data on food intake is given for comparing normal and experimental animals, the weight gains seem to be comparable, implying that the experimental animals did not voluntarily restrict their diet as has been observed in other studies. Nevertheless, paired-feeding would again have given additional validity to the study. It is of some interest that the authors noted grand-mal seizures in some of their animals (Waisman & Harlow, 1963), since about 25% of human

PKUs also exhibit such seizures (Jervis, 1963).

Cadell, Harlow, and Waisman (1962) reported EEG changes in these monkeys raised on an *L*-phenylalanine diet. Resting records of phenylalanine animals showed a predominance of slow waves, but none showed the 3/sec spikes and waves reported by Fois, Rosenberg, and Gibbs (1955) for human PKU. Normal controls showed predominance of high frequency activity. Some phenylalanine animals had reduced following responses to intermittent photic stimulation. Thus there was little similarity to EEG activity of human PKU. It would also be important to know if these monkeys *and controls* were raised on surrogate mothers. EEG changes toward slowing have been reported for short-term sensory isolation even in adult humans (Zubek, Welch, & Saunders, 1963).

Waisman, Hable, Wang, and Akert (1964) have also reported ultrastructural changes in glia cells and myelinization of brains of rats and one monkey raised on an excess phenylalanine diet. Since previous reports from this group have not adequately demonstrated PKU in rats or monkeys analogous to human PKU, these results must remain only highly suggestive.

While these studies seem the most promising, the small *N*, large variability, lack of pair-fed controls, and the possible contamination by the rearing conditions requires further study before concluding that PKU can be induced in monkeys.[5]

PKU Criteria

The disease was originally called phenylpyruvic oligophrenia in recognition of both the biochemical and behavioral characteristics. This has changed in recent years to phenylketonuria, reflecting the biochemical emphasis.[6] While biochemical investigation has grown rapidly, behavioral investigation of PKU has been neglected, consisting almost solely of intelligence testing and clinical observation. There has been no experimental study of PKU abilities or behavior. Nevertheless, it is obvious that PKU as a meaningful pathology entails a relatively specific biochemical syndrome accompanied by a rather gross behavioral syndrome labeled mental retardation. One without the other cannot be PKU in any meaningful sense. Therefore, experimental attempts to produce PKU must use both adequate biochemical criteria and behavioral criteria if they are to speak of their results as PKU. There has been no study to date meeting this requirement. Many studies have, however, assumed adequate demonstration of PKU in animals submitted to certain procedures and have consequently accepted these procedures, thereby perpetuating and compounding the error.

Chemical Criteria

Since the original discovery of phenylpyruvic acid in the urine and increased plasma phenylalanine, followed by the finding of insufficient liver phenylalanine hydroxylase activity, a number of additional biochemical differences between PKU and normal humans have been found. It is necessary that experimental attempts to produce PKU increasingly strive to duplicate whatever biochemical differences are essential and sufficient for the condition. Presently these would seem to be the presence of both phenylpyruvic acid in the urine and a serum phenylalanine level above about 20 mg.% in conjunction with reduced phenylalanine hydroxylase activity. In addition, serum tyrosine levels should be low, a result usually not found in the experimental studies thus far. It is also apparent that a general abnormality of tryptophan metabolism exists which may have a broad effect on brain development since it is a precursor of serotonin and even nicotinic acid, both of which have profound effects on the CNS and mental functioning. In this respect, a number of studies cited have demonstrated that excess phenylalanine results in decreased levels of serotonin in the brain or other tissues (see also, e.g., Culley, Saunders, Mertz, & Jolly, 1962; Hsia, Nishimura, & Brenchley, 1963).

---

[5] A more complete report of this work has since been published (see Waisman & Harlow, 1965). It is doubtful if it changes the evaluation.

[6] While the two terms are used interchangeably, it is possible that they should be reserved for separate entities in order to avoid confusion, that is, phenylpyruvic oligophrenia for the inclusive biochemical and behavioral syndrome; phenylketonuria for the biochemical symptoms without the behavioral symptoms.

It is possible that other compounds such as orthohydroxyphenylacetic acid, phenylacetic acid, phenyllactic acid, and even plasma amino acid patterns may be of direct importance for the pathology, in which case they should be included as indicators in experimental studies. Recently Boggs, McLay, Kappy, and Waisman (1963) have demonstrated another biochemical similarity of the experimental condition to that found in PKU, namely, decreased excretion of 5-hydroxyindolylacetic acid in monkeys.

Some experimenters feel that the sine qua non of phenylketonuria is the presence of phenylpyruvic acid in the urine. While this was the first recognized index giving the disease its name, it is doubtful if it should be considered the most basic characteristic of the larger entity, especially in the case of administering *D*-phenylalanine. It seems essential to use the *L*-isomer since the *D*-isomer is a nonphysiological form and is not involved in human PKU. Consequently, it can be metabolized differently—directly to phenylpyruvate by the kidney—giving positive urine tests without yielding the other metabolic derivatives possibly involved in PKU.

The picture seems further complicated by recent reports of individuals with the biochemical syndrome but normal intelligence as well as non-PKU children from PKU mothers (and other variant combinations). These would indicate that the presently implicated biochemical abnormalities and genetic link are not sufficient for understanding the basic lesion. Bessman (1964) has called attention to the possibility that metabolic abnormalities of an essential enzyme cofactor, pteridine (Kaufman, 1963), may be responsible for cases presenting unusual syndromes, implying that PKU may not be a homogeneous disease caused by a single defect.

Behavioral Criteria

Since there has been little behavioral research on PKU other than IQ testing and clinical observation, there are no specific behavioral indices whose analogs could be used in animal work. With only the rather nebulous description of "mental deficiency" to go by, the task in animal experimentation is far from simple. It is of some help to note the clinical descriptions of PKU behavior such as hyperactive, irritable, unruly, schizophrenic-like, and the existence of seizures. These may be taken as possible guides, but they are present only in some cases and are also known to be present in many other pathological conditions.

If "mental deficiency" is taken to mean impaired learning performance and adaptive behavior (social and environmental), its analog in animals should be measurable. Work to date has generally used performance on a single task (maze performance, discrimination learning, etc.) to determine "mental deficiency" with the implicit assumption that performance on other tasks will be similar (Waisman et al., 1962, is a notable exception here). This is unacceptable even by IQ-test procedure where a range of behavior is tapped on a number of subtests. This problem is a recurring one when assessing general ability in animals (cf. Rosenzweig, Krech, & Bennett, 1960; Searle, 1949, on maze "bright" and maze "dull" rats). It is possible that a concept of a general intelligence factor is less applicable to animal behavior than to human behavior.

With no guide to the type of tasks that might be discriminative for PKU individuals, it is necessary to start with a variety of traditional measures in animal behavior until appropriate information from human PKU is obtained. Certain tasks seem to have obvious worth in this connection: the Hebb-Williams mazes, discrimination learning and reversal, and avoidance learning, to cite a few. Any tasks chosen must also be regarded in relation to the methodological problems discussed below.

METHODOLOGICAL PROBLEMS

Many of the studies reviewed are permeated by methodological problems. The present section attempts to point out some of the more basic of these but makes no attempt to be comprehensive. The four sections are obviously interdependent, needing more or less emphasis according to the study considered.

Adequate Controls

Only recently have basic controls such as paired-feeding begun to be used (cf. Polidora et al., 1963) with no behavioral results as yet published. There is some question if paired feeding is a sufficient control, since the amino acid imbalance in the phenylalanine-fed group is not present in the pair-fed group. It may be better to produce a comparable amino acid imbalance by feeding a different amino acid not associated with mental deficiency.

Studies using a particularly traumatic means of induction (interperitoneal injection in neonates or very high dietary levels in gestating or lactating females, etc.) also require consideration of the selective factor of the high mortality which occurs.

Further controls of environmental-developmental effects need to be used in light of the reported effects on learning ability and brain chemistry (Hunt, 1961; Krech, Rosenzweig, & Bennett, 1960). There is too little known of the complex interaction of development and experience on later learning behavior, while the known effects on social behavior suggest that isolation-reared animals are more closely analogous to mental defectives than to normals. It is questionable, therefore, if isolation-reared animals should be used to investigate the effects of a third factor with as diffuse effects as PKU.

In studies using a dietary method of induction there has been little concern for precise control of phenylalanine content of the stock feed generally used. This is known to vary considerably and may be responsible for some of the variations obtained in results.

Lastly, there has been little replication of the biochemical criteria and less replication of the reported behavioral changes. Replication and increased numbers of animals are a necessity to discount random procedural and biological variation.

Motivation

The assessment of behavioral changes in studies of this nature is complicated by motivational differences between the groups. It is well known that amino acid imbalance produces changes in dietary habits and, specifically, that excess phenylalanine produces a voluntary decrease in food intake. Thus, if hunger is used as a motivator, the control and experimental groups will be differentially motivated. Pair feeding while controlling the effects of reduced food intake introduces its own effects of increased motivation (food-seeking) since the pair-fed group essentially is on a deprivation schedule. It is reasonable to assume that thirst is also subject to this reasoning on the basis of the interaction between food and water intake. Many studies have consequently used some sort of avoidance or escape task to circumvent these problems. The Wisconsin group's use of escape from water in maze performance (Polidora et al., 1963), and Woolley and Hoeven's (1963) use of escape from light offer promise at present. In addition, shock avoidance in a discrimination task or shuttlebox may be productive. Such avoidance motivation may not be without its own problems, for example, escape from light may be increased by heightened skin sensitivity from excess phenylalanine known to decrease pigmentation and to give rise to eczematous conditions, or escape from shock may be confounded by skin-resistance changes produced by unbalanced diets affecting the skin. These problems are avoided by allowing sufficient time to elapse after administration of excess phenylalanine so that such motivational changes can return to normal (cf. Perez, 1964). Human PKU, if not given therapy shortly after birth, results in permanent damage. There is, therefore, every reason to expect behavioral effects in experimental animals some months after cessation of phenylalanine excess if a true analog of PKU has been produced.

Developmental Variables

Since the specific lesion resulting in PKU occurs in early infancy when the brain is still developing and therapy is most effective if begun at this time, developmental factors are of utmost importance in attempts to produce PKU in animals. Many studies have begun with adolescent or young adult animals in which it is doubtful at the outset if analogous brain effects can be produced. These developmental differences seem responsible for some of the divergent results reported. The use of fetal or neonatal animals would seem necessary, but gives rise to problems of inducing PKU. Four methods have been used to date: special diet to the pregnant or lactating female, special diet to the neonate, intraperitoneal injection of the neonate, or selective breeding. While the special diet to the pregnant or lactating female is most convenient, mortality of the young is sometimes high.

The fact that the enzyme phenylalanine hydroxylase becomes active, with some variation, shortly after birth (Freedland, Krakowski, & Waisman, 1962; Kenny, Reem, & Kretchmer, 1958) should be emphasized in selecting the developmental stage with which to work.

Species Differences

Results must also be interpreted in relation to the specie upon which they have been obtained. It is obvious that the behavioral tasks chosen may have differing importance for the specie involved, for example, simultaneous versus successive discrimination. For work with monkeys, the Wisconsin group (cf. Waisman et al., 1962) has chosen a reasonable range of tasks. Just as obvious, however, is the importance of species (as well as strain) differences in biochemical lesion, for example, phenylacetic acid metabolism. Bessman (1964) has recently pointed out that biochemical differences may exist phylogenetically which may "render inapplicable to man some of the observations on experimental PKU in the rat and monkey . . . ." Rats, for example, have 10 to 15 times more phenylalanine hydroxylase activity than do humans (Waisman & Harlow, 1963), making it exceedingly difficult to produce an excess of phenylalanine, while monkeys have levels comparable to humans. Upon these considerations, primates would seem to be the best suited for this work.

GENERAL CONCLUSIONS

The production of mental retardation in experimental animals by biochemical, environmental, or other experimental means provides an important tool for the study of intellectual ability. It is doubtful, however, if this has yet been realized for PKU. It seems that claims for the production of PKU in animals have been premature. There have been no studies meeting the criteria discussed and also free of basic methodological errors. While many have demonstrated an increase in serum phenylalanine, it is generally far below the levels found in human PKU even though phenylalanine hydroxylase activity may be reduced 20–40%. Similarly, many studies have reported behavioral effects whose validity and reliability are questionable. Some laboratories have now apparently begun to systematically investigate the conditions needed to adequately and reliably produce the biochemical criteria and the role of developmental variables. The coupling of these endeavors with more adequate behavioral assessment should produce the information now lacking in this potentially fruitful area.

Many studies have observed severe growth inhibition in their experimental animals. There is little evidence for reduced body size or weight in human PKU (Jervis, 1954, 1963) which may lead one to reason that the growth inhibition and mortality observed in animal studies is atypical and may reflect processes different than true PKU but which could easily result in performance decrements. These processes may involve simple malnutrition due to reduced dietary intake and amino acid imbalance. The techniques used to produce excess phenylalanine require that it has to be absorbed into the blood from the gut or peritoneal cavity with the possibility of inhibitory effects of phenylalanine on active transport of other amino acids which may also produce a malnutrition state. In human PKU, by contrast, the excess is produced directly in the blood via the liver. There is evidence that malnutrition at certain stages of development can adversely affect behavior. In this connection kwashiorkor, a general protein deficiency widespread in South American and African-Asian countries and which has detrimental results on intellectual and behavioral development (cf. Scrimshaw & Behar, 1961), may be of some relevance. The effects of this deficiency are, however, possibly reversible in contrast to PKU. It is of interest that some liver enzyme activities are lost early in dietary protein deficiency (Knox, Auerbach, & Linn, 1956). Recently, Cowley and Griesel (1959, 1962, 1963, 1964) have reported behavioral deficits in first- and second-generation rats raised on low protein diets.

Many of the PKU studies to date are preliminary reports and convention papers which rely upon unpublished data and observation. In the future it is to be hoped that more extensive and complete reports will appear so that results can be more readily evaluated. Many problems in this work are similar to those encountered in experimental behavior genetics. Thus, the investigator attempting to produce PKU would do well to be cognizant of that work. Lastly, the studies reviewed offer the promise of experimental production of PKU, an important step in our knowledge of the relations between biochemistry, development, and intelligence, as well as a needed change in orientation of research on mental retardation.

REFERENCES

Ammon, R. Versuche zur Phenylketonurie beim Kaninchen. *Zeitschrift fur physiologische Chemi,* 1961, 324, 122–124.

Auerbach, V. A., Waisman, H. A., & Wyckoff, L. B. Phenylketonuria in the rat associated with decreased temporal discrimination learning. *Nature,* 1958, **182**, 871–872.

Bessman, S. P. Some biochemical lessons to be learned from phenylketonuria. *Journal of Pediatrics,* 1964, **64**, 828–838.

Boggs, D. E., McLay, D., Kappy, M., & Waisman, H. A. Excretion of indolyl acids in phenylketonuric monkeys. *Nature,* 1963, **200**, 76.

Boggs, D. E., & Polidora, V. J. Biochemical and behavioral observations of phenylketonuria (PKU) in rats. *Federation Proceedings,* 1963, **22**, 515. (Abstract)

Boggs, D. E., Polidora, V. J., & Waisman, H. A. Further attempts to induce PKU in rats. *Federation Proceedings,* 1962, **21**, 3. (Abstract)

Boggs, D. E., Rosenberg, R., & Waisman, H. A. Effects of phenylalanine, phenylacetic acid, tyrosine, and valine on brain and liver serotonin in rats. *Proceedings of the Society for Experimental Biology and Medicine,* 1963, **114**, 356–358.

Boggs, D. E., & Waisman, H. A. Effects on the offspring of female rats fed phenylalanine. *Life Sciences,* 1962, **8**, 373–376.

Boggs, D. E., & Waisman, H. A. Influence of excess dietary phenylalanine on pregnant rats and their fetuses. *Proceedings of the Society for Experimental Biology and Medicine,* 1964, **115**, 407–410. (a)

Boggs, D. E., & Waisman, H. A. Biochemical correlates in rats with phenylketonuria. *Archives of Biochemistry and Biophysics,* 1964, **106**, 307–311. (b)

Cadell, T. E., Harlow, H. F., & Waisman, H. A. EEG changes in experimental phenylketonuria. *Electroencephalography and Clinical Neurophysiology,* 1962, **14**, 540–543.

Coleman, D. L. Phenylalanine hydroxylase activity in dilute and nondilute strains of mice. *Archives of Biochemistry and Biophysics,* 1960, **91**, 300.

Cowley, J. J., & Griesel, R. D. Some effects of a low protein diet on a first filial generation of white rats. *Journal of Genetic Psychology,* 1959, **95**, 187–201.

Cowley, J. J., & Griesel, R. D. Pre- and postnatal effects of a low protein diet on the behaviour of the white rat. *Psychologia Africana,* 1962, **9**, 216–225.

Cowley, J. J., & Griesel, R. D. The development of second-generation low-protein rats. *Journal of Genetic Psychology,* 1963, **103**, 233–242.

Cowley, J. J., & Griesel, R. D. Low protein diet and emotionality in the albino rat. *Journal of Genetic Psychology,* 1964, **104**, 89–98.

Culley, W. J., Saunders, R. N., Mertz, E. T., & Jolly, D. H. Effect of phenylalanine and its metabolites on the brain serotonin level of the rat. *Proceedings of the Society for Experimental Biology and Medicine,* 1962, **111**, 444–446.

Fois, A., Rosenberg, C., & Gibbs, F. A. The electroencephalogram in phenylpyruvic oligophrenia. *Electroencephalography and Clinical Neurophysiology,* 1955, **7**, 569–572.

Freedland, R. A., Krakowski, M. C., & Waisman, H. A. Effect of age, sex and nutrition on liver phenylalanine hydroxylase activity in rats. *American Journal of Physiology,* 1962, **202**, 145–148.

Green, P. C., & Gordon, M. Maternal deprivation: Its influence on exploration in infant monkeys. *Science,* 1964, **145**, 292.

Harlow, H. F. The nature of love. *American Psychologist,* 1958, **13**, 673–685.

Harlow, H. F. The development of learning in the Rhesus monkey. *American Scientist,* 1959, **47**, 459–479.

Harlow, H. F. The heterosexual affectional system in monkeys. *American Psychologist,* 1962, **17**, 1–9.

Hsia, D. Y-Y., Nishimura, K., & Brenchley, Y. Mechanisms for the decrease of brain serotonin. *Nature,* 1963, **200**, 578.

Huang, I., Tannenbaum, S., Blume, L., & Hsia, D. Y-Y. Metabolism of 5-hydroxyindole compounds in experimentally produced phenylketonuric rats. *Proceedings of the Society for Experimental Biology and Medicine,* 1961, **106**, 533.

Hunt, J. McV. *Intelligence and experience.* New York: Ronald Press, 1961.

Jervis, G. A. Phenylpyruvic oligophrenia (phenylketonuria). *Proceedings of the Association for Research on Nervous and Mental Disease,* 1954, **33**, 259–282.

Jervis, G. A. The clinical picture. In F. L. Lyman (Ed.), *Phenylketonuria.* Springfield, Ill.: Charles C Thomas, 1963.

Kaufman, S. The structure of the phenylalanine hydroxylation co-factor. *Proceedings of the National Academy of Sciences,* 1963, **50**, 1085–1093.

Kenney, F. T., Reem, G. H., Kretchmer, N. Development of phenylalanine hydroxylase in mammalian liver. *Science,* 1958, **127**, 86.

Knox, W. E. Incidence and inheritance. In F. L. Lyman (Ed.), *Phenylketonuria.* Springfield, Ill.: Charles C Thomas, 1963.

Knox, W. E., Auerbach, V. H., & Linn, E. C. Enzymatic and metabolic adaptations in animals. *Physiological Review,* 1956, **36**, 164–254.

Kretch, D., Rosenzweig, M. R., & Bennett, E. L. Effects of environmental complexity and training on brain chemistry. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 509–519.

Loo, Y. H., Diller, E., & Owen, J. E. Effect of phenylalanine diet on learning in the rat. *Nature,* 1962, **194**, 1286.

Louttit, R. T. Effect of phenylalanine and isocarboxazid feeding on brain serotonin and learning behavior in the rat. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 425–428.

McKean, C. M., Schanberg, S. M., & Giarman, N. J. A mechanism of the indole defect in experimental phenylketonuria. *Science,* 1962, **137**, 604–605.

McMillan, D. E. Learning and performance: A distinction. *Science,* 1964, **145**, 952.

Neill, K. C., & Langford, H. G. Studies of the production of phenylketonuria in the rat. *Federation Proceedings,* 1961, **20**, 2. (Abstract)

Perez, V. J. Phenylpyruvic oligophrenia in the rat. *American Psychologist,* 1964, **19**, 498. (Abstract)

Perez, V. J., & Schmidt, H. Experimental phenylketonuria: Effects of excessive amount of L-phenylalanine upon maze learning ability of the albino rat. *American Psychologist,* 1963, 18, 430. (Abstract)

Polidora, V. J., Boggs, D. E., & Waisman, H. A. A behavioral deficit associated with phenylketonuria in rats. *Proceedings of the Society for Experimental Biology and Medicine,* 1963, **113**, 817–820.

Rosenzweig, M. R., Krech, D., & Bennett, E. L. A search for the relations between biochemistry and behavior. *Psychological Bulletin,* 1960, **57**, 476–492.

Schanberg, S. M., McKean, C. M., & Giarman, N. J. A mechanism of the indole defect in experimental phenylketonuria. *Federation Proceedings,* 1962, **21**, 269. (Abstract)

Schrimshaw, N. S., & Behar, M. Protein malnutrition in young children. *Science,* 1961, **133**, 2039–2047.

Searle, L. V. The organization of hereditary maze brightness and maze dullness. *Genetic Psychology Monographs,* 1949, **39**, 279–325.

Waisman, H. A. Advances in the understanding of the metabolic abnormalities in phenylketonuria. In P. W. Bowman & H. V. Mautner (Eds.), *Mental retardation: Proceedings of the First International Conference on Mental Retardation.* New York: Grune & Stratton, 1960.

Waisman, H. A. Role of phenylalanine excess in experimental phenylketonuria. In R. Korey et al. (Eds.), *Ultrastructure and metabolism of the nervous system.* Vol. 50. *Research publications of the Association for Research on Nervous and Mental Disease.* Baltimore: Williams & Wilkins, 1962.

Waisman, H. A. The experimental approach. In F. L. Lyman (Ed.), *Phenylketonuria.* Springfield, Ill.: Charles C Thomas, 1963.

Waisman, H. A., Hable, K., Wang, H. L., & Akert, K. Some ultrastructural changes in the brain of phenylketonuric rats and monkeys. In W. A. Himwich & H. E. Himwich (Eds.), *The developing brain: Progress in brain research.* Vol. 9. New York: Elsevier, 1964.

Waisman, H. A., & Harlow, H. F. The biochemical induction of phenylketonuria in monkeys and rats. In, *Proceedings of the Second International Congress on Mental Retardation: 1961.* Part 1. Basel: S. Karger, 1963.

Waisman, H. A., & Harlow, H. F. Experimental phenylketonuria in infant monkeys. *Science,* 1965, **147**, 685–695.

Waisman, H. A., Wang, H. L., Palmer, G., & Harlow, H. F. Phenylketonuria in infant monkeys. *Nature,* 1960, **188**, 1124–1125.

Waisman, H. A., Wang, H., Palmer, G., & Harlow, H. F. Experimental phenylketonuria in infant monkeys. In, *Proceedings of the London Conference on the Scientific Study of Mental Retardation, July, 1960.* Dagenham, England: May & Baker, 1962.

Wang, H. L., Harwalkar, V. H., & Waisman, H. A. Effect of dietary phenylalanine and tryptophan on brain serotonin. *Archives of Biochemistry and Biophysics,* 1962, **97**, 181–184.

Wang, H. L., & Waisman, H. A. Experimental phenylketonuria in rats. *Proceedings of the Society for Experimental Biology and Medicine,* 1961, **108**, 332–335.

Wassmuth, M. P. Über die oligophrenia phenylpyruvica. Klinische, tierexperimentelle und biochemische, insbesondere papierchromatographische Untersuchungen. *Annalis Universitatis Saraviensis Medicinae,* 1958, **6**, 184–215.

Watt, D. D., & Vandevoorde, J. P. The inhibition of phenylalanine hydroxylase by DL-4-fluorophenylalanine. *Federation Proceedings,* 1964, **23**, 146. (Abstract)

Wimer, R., & Weller, S. Discrimination learning in the mouse. Paper presented at Eastern Psychological Association Convention, 1964.

Woolley, D. W. *The biochemical bases of psychoses.* New York: Wiley, 1962.

Woolley, D. W., & Hoeven, T. v. d. Alteration in learning ability caused by changes in cerebral serotonin and catechol amines. *Science,* 1963, **139**, 610–611.

Woolley, D. W., & Hoeven, T. v. d. Serotonin deficiency in infancy as one cause of a mental defect in phenylketonuria. *Science,* 1964, **144**, 883–884. (a)

Woolley, D. W., & Hoeven, T. v. d. Experimental production of a mental defect of phenylketonuria and its prevention by serotonin derivatives. *Federation Proceedings,* 1964, **23**, 146. (Abstract) (b)

Woolley, D. W., & Hoeven, T. v. d. Prevention of a mental defect of phenylketonuria with serotonin cogeners such as melatonin or hydroxytryptophan. *Science,* 1964, **144**, 1593–1594. (c)

Yuwiler, A., & Louttit, R. T. Effects of phenylalanine diet on brain serotonin in the rat. *Science,* 1961, **134**, 831–832.

Zubek, J. P., Welch, G., & Saunders, M. G. Electroencephalographic changes during and after 14 days of perceptual deprivation. *Science,* 1963, **139**, 490–492.

(Received November 9, 1964)

Psychological Bulletin
1965, Vol. 64, No. 1, 65–73

SOCIAL REINFORCEMENT OF CHILDREN'S MOTOR BEHAVIOR:

A REVIEW [1]

DAVID A. PARTON
Institute of Child Behavior and Development, University of Iowa
AND ALAN O. ROSS
Pittsburgh Child Guidance Center, Pennsylvania

This review highlights a number of methodological weaknesses and problems of measurement found in many studies which have attempted to evaluate the effectiveness of positive social reinforcement on repetitive motor behavior in children. The omission of control groups has frequently resulted in the confounding of reinforcer effects with regression-to-the-mean or warm-up effects. Problems relating to the use of difference scores, the effect of various schedules of reinforcement, the measurement of developmental changes, and the function of mediating response strategies are discussed. Recommendations for future research focus on the possible role of awareness of the reinforcement contingency and on the influence of other mediating responses.

Recent years have seen the publication of a number of studies attempting to influence children's motor behavior by "positive social reinforcement." In most of these studies, the subject's task was to take marbles from a bin or dispenser, one at a time, and to insert them in a hole in the apparatus; the number of holes available has ranged from two to nine. Social reinforcement has consisted chiefly of praise statements by the experimenter. In some cases the experimenter also emitted smiles and nods. This review will focus on apparent deficencies in methodology or design in studies dealing with the influence of verbal reinforcement on children's repetitive motor-task performance of the type described. Each of the empirical indices used to measure social-reinforcer effectiveness will be considered, followed by a discussion of some problems of method.

PREFERENCE-RATIO MEASURES

The measurement of the presumed effect of social reinforcement has been expressed in terms of a change in ratio of critical over total responses before and after introduction of the reinforcement condition. The "preference ratio," introduced by Gewirtz and Baer (1958b), represents the number of responses to the less-preferred hole (X_b) divided by the total number of responses to both holes ($X_b + Y_b$) during the last minute of a 4-minute nonreinforcement base-line period. A comparable preference ratio ($X_r/X_r + Y_r$) is obtained for the reinforcement period or for several intervals within this period. In this ratio X_r represents the number of responses to the hole which had been less preferred during the base-line period, and is thus the hole to which responses were reinforced during the experimental period. The difference between the values, ($X_r/X_r + Y_r$) and ($X_b/X_b + Y_b$), is a preference-change score. In some instances the score calculated for a subject has been a mean or median value of several preference-change scores computed by subtracting a single base-line-period ratio from the ratio obtained for each of several consecutive reinforcement intervals.

The Gewirtz and Baer preference change score (1958a, 1958b) has also been used by Walters and Foote (1962), Walters and Ray (1960), and Zigler and Kanzer (1962). Unfortunately this scoring system entails the questionable assumption that the base-period

---

[1] Preparation of this review was supported in part by Grant MH588 from the National Institute of Mental Health, United States Public Health Service. The writers are indebted to D. M. Baer, L. D. DeNike, J. R. Patterson, H. W. Stevenson, and E. Zigler for helpful criticism of an earlier version of this paper, but the opinions herein expressed are solely the responsibility of the authors.

preference becomes stable after a short interval of responding, and can be reliably measured by a ratio based on the subject's fourth minute of performance on the task. Since none of these studies used nonreinforced control groups, the reported effects of social reinforcement may have been confounded with the statistical artifact of regression to the mean, with trends within time, or both.[2]

Regression to the Mean

Regression to the mean is found when later measures are taken on groups of subjects originally constituted on the basis of their extreme scores. In the case of preference-change scores for the two-hole, marble-dropping task, the regression effect could stem from the procedure of selecting the subject's less-preferred position (hole) response as the one to be reinforced. In a study by Patterson and Hinsey (1964), for example, some subjects exhibited no stable position preference during the base period but fluctuated around the chance (50%) point between the two possible response alternatives. For such subjects the use of a brief base period for the determination of the base-period preference score, $X_b/X_b + Y_b$, should, even without reinforcement, tend to yield $X_r/X_r + Y_r$ values larger (i.e., closer to chance level) than the base-period preference scores.

Cairns (1962), using a three-response task, has demonstrated that preference-change scores can reflect such a regression effect. In this study, subjects worked a pinball machine which could be operated by any one of three distinctly colored knobs. During the base period the mean preference ratio, $X_b/X_b + Y_b + Z_b$, of both the selectively reinforced and the randomly reinforced control groups was approximately .23. Following the base period the selectively reinforced group received a praise statement every time the least-preferred response occurred, while the control group was praised for every third response regardless of which knob was being used. During the first 10 trials of the reinforcement period both groups rose to a mean preference ratio of approximately .33, a value which corresponds to chance performance. It would seem probable that the preference ratio increases of .10 (.33–.23), found irrespective of selective reinforcement, were due to regression to the mean.

The importance of this regression artifact for the two-hole, marble-dropping task has not been explored. The effect appears to be inconsequential, however, when the preference ratio is derived from at least 4 minutes of performance (Patterson & Hinsey, 1964).

Some investigators have avoided the possible influence of statistical regression. McCullers and Stevenson (1960), for example, randomly selected the response alternative to be reinforced on a task which involved pushing one of three knobs on each trial. They report no significant differences between reinforcement and control groups of 8–10 year old subjects.

As yet untried is the possibility of avoiding the regression artifact by the use of a double base-line period. The first interval of the base period could be used solely to determine the subject's less-preferred response. The second interval could then be used to derive the base-period preference ratio. While this value would be independent of statistical regression, a control group should be employed to ascertain whether systematic trends occur over time.

Stability of Individual Differences

Investigators have tried to relate preference-change scores to measures of child personality or parental child-rearing behavior (Cairns, 1962; Patterson, Littman, & Hinsey, 1964), but in the absence of substantial relationships with other variables the reliability of preference change as an index of inter-individual differences should be established.

Patterson and Hinsey (1964) report a test-retest correlation for preference-change scores of .75 ($N = 24$) when testing sessions were separated by 1 week. This value is surprisingly high for the reliability of a change score (see Bereiter, 1963) and should be evaluated in terms of certain procedural details. Praise statements were dispensed on a 1:1 ratio schedule for the first 20 less-preferred responses and then on a 2:1 ratio schedule for the next 40 appropriate responses. This schedule made it possible for the subject to develop a correct performance strategy unrelated to the reinforcement contingency. By increasing response rate, subjects were able to increase the number of reinforcements obtained per unit time without changing the proportion of responding to the two holes.[3] Adoption of this strategy, or of an alternation strategy, would result in relatively small but comparable preference-change scores during both the first and second sessions. The reliability coefficient of .75 reported by Patterson et al. (1964) would probably have been lower had the experimental task permitted only one strategy for increasing the number of reinforcements per unit time.

---

[2] Reference to any study in connection with a method problem should not be taken to imply that the results of the study are necessarily rendered ambiguous, since in some of these investigations the manipulation of an independent variable provided confirmation of the empirical dependent variable.

In addition to the point just discussed, there exists the possibility that subjects who are aware of the reinforcement contingency during the first session may enter the base period of the second session with expectancies or sets affecting their performance in diverse ways. It thus appears that the test-retest correlation may be an inappropriate statistic for evaluating the reliability of preference-change scores. An alternative approach lies in estimation of the reliability of preference-change scores from the internal reliability of performance during the first-base and first-conditioning periods and from the correlation between performance during these two periods (Guilford, 1954, p. 393).

RATE OF RESPONSE MEASURES

Sometimes social reinforcement is studied by observing whether the subject speeds up after introduction of the reinforcer. Rate-increase scores based on this approach are typically difference scores derived by subtracting the response rate during a base period from the response rate during an experimental period.

Some experimenters implicitly make the crucial but questionable assumption that in the absence of reinforcement the response rate of child subjects is stable over time. This assumption has led to studies lacking nonreinforced control groups because each subject is thought to be "his own control" (Grossman, 1963; Stevenson, 1961; Stevenson, Keen, & Knights, 1963; Stevenson & Knights, 1962b; Stevenson & Odom, 1962). It is unlikely, however, that measures based on brief base periods can serve as controls for time-related changes such as warm-up effects, fatigue effects, and the effects of subjects' varying hypotheses and motivational states.

The existence of such time-related changes was reported by Zigler (1963). Since this study did not incorporate a base period, the responses during the first minute of the 6-minute reinforcement period will be compared with later performance, for the purpose of this exposition. The nonreinforced group ($N = 64$) showed a mean-per-minute rate increase identical with that of the reinforced group ($N = 64$)—1.6 responses. While this is a relatively small upward drift in response rate, its importance can be seen in relation to a study by Stevenson and Odom (1962). There the overall, mean-per-minute rate increase for subjects receiving social reinforcement was less than 1.0.

Suggestive evidence for the existence of an upward response-rate drift can be found in a study by Stevenson and Knights (1962a) who report an increase in response rate over three separate testing sessions. Of tangential interest is the failure to demonstrate the influence of positive social reinforcement when a nonreinforcement control group was employed (Kelly & Stevens, 1964;[4] Parton & Ross, 1964; Stevenson & Fahel, 1961; Zigler, 1963). It should be noted, however, that Andronico (1963) reports a significantly higher, lever-pulling rate during a 7-minute reinforcement period for a reinforced group than for a nonreinforced control group.

Several factors may contribute to response-rate drift. The most fundamental of these is probably the warm-up effect which is frequently encountered during the initial periods of motor-task performance; another may have to do with the availability of a terminal goal. In studies where subjects are provided with a supply of marbles to be dropped in a hole (e.g., Stevenson & Fahel, 1961) the presence of the supply box furnishes a clear terminal goal, that of emptying the box. The fact that the box may hold an abundant supply of marbles (as many as 2,200 in the study by Stevenson and Fahel) may have little influence on young children who might adopt the unrealistic goal of emptying the box and, in working toward this goal, change their response rate.

---

[3] While this behavior certainly represents a response to reinforcement, it is not reflected by the index of reinforcer effectiveness and, in this context, must be considered to be an artifact.

[4] Analysis performed by the reviewers using the original data presented by Kelly (1962).

It is possible that many children in psychological experiments expect feedback on their performance from the adult experimenter. When the experimenter remains silent, as he does when running a nonreinforced control group, a child subject may institute a strategy of increasing his rate of response in order to assure himself that the experimenter is still aware of him, or to elicit from the adult a comment which might indicate that he is doing satisfactorily or has reached the stopping criterion. Should such an effect operate with a nonreinforced control group a measure based on rate of response would, of course, obscure the social-reinforcement effect.

In addition to the two aforementioned problems which beset measures based on the rate of response, there is a question about the stability of rate-change scores. Patterson and Hinsey (1964) had the same female adult administer social reinforcement to 24 child subjects for two sessions 1 week apart. They report[5] a test-retest correlation of .19 which might be interpreted as reflecting low stability of the rate-increase measure. The authors note, however, that during the second session the range of rate increase scores was severely constricted. If social reinforcement is effective in increasing the performance rate of children during the first experimental period, their high response rate might still be in evidence during the base period of the second session. The authors conclude that this increment in response rate during the base period of the second session constricted the range of possible rate-increase scores, reducing the correlation between the two testing sessions, and thus produced a deceptively low index of reliability. The suggestion offered above for evaluating the stability of preference change scores is also relevant for rate-change scores.

TIME MEASURES

Studies using time measures to evaluate the effectiveness of social reinforcement have all included nonreinforced control groups, thus providing a clear test of the influence of the experimental manipulation.

One such measure is the total time the child spends at the task. This approach was used by Zigler, Hodgden, and Stevenson (1958); Zigler and Williams (1963); and Shepps and Zigler (1962); all of whom report significant reinforcement effects with samples of retarded children. On the other hand, Green and Zigler (1962) and Zigler (1961) found no difference between reinforced and nonreinforced groups for samples of retarded children. Two studies report nonsignificant reinforcement effects with normal children (Green & Zigler, 1962; Zigler, Hodgden, & Stevenson, 1958) while more recently Zigler (1963) found a significant effect with samples of normal and retarded children. In summary, there appears to be evidence for the effectiveness of social reinforcement on time at the task with both normal and retarded children.

Some studies have used the number of marbles inserted as a measure of performance. Such a score is the product of response rate and the length of time the subject stays at the task. To date there has been one demonstration that social reinforcement affects this measure of test performance. Stevenson and Cruse (1961) found a significant overall difference in the performance of a reinforced group, an experimenter-present control group, and an experimenter-absent control group. The mean number of marbles dropped by the reinforced group was more than four times that of the experimenter-present control group, presumably a significant difference. Stevenson and Snyder (1960) report a mean number of 140.1 marbles inserted by an experimenter-present control group while the reinforced group inserted a mean of only 126.3 marbles, a nonsignificant difference.

---

[5] In an unpublished, expanded version of Patterson, Littman, and Hinsey (1964).

USE OF DIFFERENCE SCORES

When base performance is correlated with gain scores the determinants of base performance become determinants of change. Correlations greater than −.30 between base rate and rate change have been noted in three of the four studies reporting this correlation (Patterson & Hinsey, 1964; Stevenson, 1961; Stevenson & Fahel, 1961; Zigler, 1963).[6] For preference ratios, Patterson and Anderson (1964) report negative correlations ranging from −.09 to −.80 for several studies. Since analyses based on difference scores fail to identify the source of variation producing an effect, a more appropriate procedure can be provided with analysis of variance. Base-period and experimental-period scores can be entered as two subgroups in a within-subjects factor, with the experimental groups constituting a between-subjects factor (e.g., see Winer, 1962, p. 302 ff.). A significant interaction between these two factors is equivalent to a significant group effect based on difference scores. In addition, this analysis-of-variance design provides information indicating which relationship(s) produced the significant interaction by inspection of the cell means.[7]

The aforementioned factorial design appears more appropriate than an analysis-of-covariance procedure in which experimental and base-period performance serve as the criterion and concomitant variables, respectively. The relatively low reliability of base-period rate and preference scores (Patterson & Hinsey, 1964; Stevenson & Knights, 1962a) indicates that analysis of covariance would fail adequately to correct for initial differences. Furthermore, for the studies under review the data may not have met the analysis of covariance assumption of linear regression of the criterion measures on the concomitant measures (Lindquist, 1953, p. 323). The previously noted negative correlations between base and change performance probably reflect a ceiling effect (at least for the rate measure). This ceiling effect would be reflected in a departure from linearity of regression of the criterion measures on the concomitant measures.

Some investigators have exercised caution in using difference scores as an interindividual index of social-reinforcement responsiveness. In two studies (Patterson & Anderson, 1964; Patterson, Littman, & Hinsey, 1964) an attempt was made to eliminate the influence of magnitude of base-period preference (based on an N of 160). If the usual assumption is made that error of measurement is not correlated with magnitude of the control variable (base-period preference ratios) a problem is posed by the fact that those subjects who exhibit base-period preference ratios close to .50 also tend to show small preference-change scores (Patterson & Hinsey, 1964). For base preferences nearest to .50, subjects' change scores will thus incorporate a greater proportion of measurement error than will the change scores of subjects with more extreme base-period preferences. The T transformation, however, serves to equate extreme scores at each level without regard to this proportional difference in error of measurement, thereby resulting in unreliable change scores for subjects at the level nearest .50. Considering only the reliability problem and the necessity of a large sample for deriving T-score values this technique cannot be recommended.

There appears to be little, if any, justification for the use of difference scores in investigating relationships among individual difference variables. Such analysis fails to indicate whether the correlations obtained are a function of base-period relationships, experimental-period relationships, or differences in variability between the two testings. The preferred alternative is to analyze base and experimental performance separately, and to investigate differential gains in performance by computing partial correlations between experimental-period performance and each individual difference variable, with base-period performance as the control variable (see Lord, 1963).

SCHEDULES OF REINFORCEMENT

A variety of schedules of reinforcement have been used in the studies under consideration.

---

[6] All of these negative correlations were spuriously inflated because the same errors of measurement are included in both sets of scores being correlated. Lord (1963, p. 34) offers a formula which estimates the correlations between true base rate and true change with errors of measurement partialed out.

[7] The authors are indebted to L. G. Humphreys for suggesting this mode of analysis.

While acknowledging that there are practical advantages for selecting a specific schedule (e.g., controlling the number of reinforcements per unit time) there often has been no justification for the choice of a particular schedule. Whenever interval schedules are used in conjunction with a rate of response index of reinforcement effect it is impossible for the subject to change the number of reinforcements per unit time (Hartup, 1964; Parton & Ross, 1964; Stevenson, 1961; Stevenson & Fahel, 1961; Stevenson et al., 1963; Stevenson & Knights, 1962a, 1962b). Thus, regardless of the subject's response rate, compliance with the experimenter's initial instructions to perform the task results in a fixed number of reinforcements (due to all subjects operating at the limit characteristic of the schedule). The use of a schedule which is in effect "non-rate-contingent" may entail either the assumption that reinforcement has an, as yet unspecified, energizing effect, or the belief that the differential influences produced by ratio or interval schedules with tangible reinforcers (Long, Hammack, May, & Campbell, 1958) occur upon brief exposure of a subject to a social-reinforcement situation. When the dependent variable under investigation is response rate it would seem more appropriate to use a ratio schedule of reinforcement since such a schedule makes the number of reinforcements per unit time contingent on the subject's response rate (Andronico, 1963; Lockwood, 1963; Stevenson & Odom, 1962; Zigler, 1963).

When a particular response class is being reinforced, as in the reinforcement of the least-preferred response on the two-hole, marble-dropping task, even a ratio schedule of reinforcement poses problems. This stems from the fact that a subject can increase the number of praise statements per unit time by either increasing the rate of response to both holes or changing the proportion of marbles dropped into each hole. Since Patterson et al. (1964)[5] found no correlation between rate change and preference change it is quite conceivable that some subjects may respond to the reinforcement manipulation with a rate-increase strategy while the experimenter is measuring only change in preference. It should also be considered that the use of a 100% fixed-ratio schedule, even for a short period, may have the effect of reinforcing a subject's alternation behavior rather than a position response as intended by the investigator. Some of these difficulties can be avoided if the two-hole, marble-dropping situation is viewed not as a simple operant task, but as a two-choice, decision-making task in which the subject determines the intertrial interval. If a standard intertrial interval were to be used with the two-hole, marble-dropping task it should eliminate the subject's confounding strategy of increasing his rate of response and might thus provide a more sensitive assessment of the influence of social reinforcement. In addition the use of a fixed intertrial interval can reduce the subject's response rate and thereby lower the possibility of a pattern of responses being perceived as the contingency for reinforcement.

DEVELOPMENTAL CHANGES

Careful consideration must be given to what constitutes appropriate dependent variables when assessing developmental changes in responsiveness to social reinforcement. McCullers and Stevenson (1960) found that 3–5 year old children exhibit stronger position preferences than 8–10 year olds. Two opposite but equally unjustified conclusions might be drawn from such age-related differences in position preference. One might conclude that the younger children are more responsive to social reinforcement if, due to their stronger position preference, they enter the reinforcement period with smaller base-period, preference-ratio scores, thus raising the maximum increase scores possible for them. While this effect would bias the results in favor of younger children, the opposite conclusion might result from the employment of a ratio reinforcement schedule where, because of their stronger position preference, younger children would receive fewer reinforcements per unit time.

Another stereotyped mode of responding found in young children is the tendency to alternation of response. Craig and Myers (1963) found that kindergarten children display more alternation behavior than do children in the fourth and eighth grades. Patterson and Hinsey (1964) report that 49% of those subjects who alternated during the base period showed no response to the social-reinforcement manipulation. Such age-related differences in behavior suggest that preference-change scores as currently used are not suitable for work attempting to assess developmental changes.

The use of rate-increase scores for assessment of developmental changes in response to social reinforcement also embraces measurement problems. Stevenson (1961) has reported that mean-base-rate triples from ages 3 and 4 to ages 9 and 10. This marked difference in base-rate performance suggests that ceiling effects may partially determine the results of change-score analysis for such data.

Mediating Responses

The heuristic value of a cognitive approach to social reinforcement has been established (Spielberger, 1965), and DeNike (1964) has presented evidence indicating that, at least with college subjects, awareness is a prerequisite for such learning. Among those studies pertaining to the social reinforcement of children's motor behavior, only Patterson and Anderson (1964) have investigated the role of the subject's awareness of the reinforcement contingency. Patterson and Anderson report nonsignificant correlations between awareness ratings based on postexperimental interviews and preference-change scores. The interview used to assess awareness of the reinforcement contingency was developed by Patterson et al. (1964). It was designed to force the subject who did not spontaneously verbalize a strategy to "guess" the hole on which reinforcement had been contingent. Using a 4-point scale for rating awareness of the contingency, subjects who guessed the incorrect hole were assigned a score of 1 while those who guessed the correct hole earned a score of 3 (the scores of 2 and 4 were assigned to irrelevant strategies or spontaneous verbalizations of correct strategies, respectively). This procedure, which arbitrarily weights the random guesses of some subjects high and of others low, confounds the evaluation of the influence of awareness on preference-change scores. Patterson and Hinsey (1964) have furthermore pointed out that those subjects who show frequent response alternations are reinforced for maintaining this alternation behavior and frequently show zero preference-change scores. Those subjects who alternate and are judged aware would thus tend to attenuate the correlation between awareness and preference change.

The two-hole, marble-drop task provides an excellent means of investigating the influence of both awareness of the correct strategy and the incidence and influence of other response strategies, such as a speed strategy and an alternation strategy. Subjects could be assigned to strategy groups on the basis of a postexperimental interview and their experimental performance then analyzed in terms of preference, rate of response, and response alternation. In addition, it would be important to consider the subject's verbal report of how much he liked receiving the praise words and whether he implemented his strategy (i.e., worked to maximize the amount of reinforcement).

Conclusions

Precise application of operant methodology has proven of value in the study of reinforcement effects with children (Bijou & Baer, 1963; Long et al., 1958), but many of the studies here reviewed abbreviated operant technique by using only one estimation of base-line performance and one estimation of reinforced performance. If the investigator is serious about adopting an operant approach, repeated assessments of base-line and reinforced performance should be obtained in order to show whether a reliable reinforcement effect is present. Whatever method orientation is adopted, however, it appears fruitless to use experimental situations which promote the development of response strategies unrelated to the empirical dependent variable.

There is a need for the development of more appropriate tasks for assessing the influence of social reinforcement. The assumption that tasks must have low intrinsic interest is subject to question. When intrinsic interest is low the subject's continued performance is evidently a function of the experimenter's instructions, which suggests that performance is aimed at obtaining or maintaining the experimenter's approval. To a great extent base-line performance may be contingent on the subject's desire for social approval. Under these circumstances the introduction of praise statements following a long base period may serve as a cue to terminate performance, emphasize a strategy which was started during the base period, or search for the experimenter's reinforcement contingency in order to terminate a boring task. Some of these problems could be circumvented by devising a task in which the occurrence of either of two instrumental responses results in an interesting visual display (or tangible reward). Social reinforcement could be made contingent on one of these responses and a preference measure employed as an index of reinforcer effectiveness.

Among the studies reviewed, even the investigations which report significant reinforcement effects do not reveal marked control over behavior. As Bijou and Baer (1963) have indicated, this may be due to the same stimulus serving as a positive reinforcer for one child, a neutral event for another, and possibly even as a negative reinforcer for a third. It may thus be unreasonable to expect that social reinforcement will uniformly facilitate performance in a random sample of children. It must also be acknowledged that there has been no demonstration that the stimuli introduced as positive reinforcers in these studies served more than a cue function. Might it not be enlightening to introduce a random reinforcement condition along with an adequate posttest interview with the aim of determining whether preference-change scores and the report of a least-preferred-hole strategy varied across contingent and random-reinforcement conditions?

Investigation of the events and processes which lead to individual variation in the incentive value of social reinforcement for children, while a difficult venture, holds prospects of illuminating much of the socialization process. Such endeavors deserve and demand a sound methodological base.

REFERENCES

Andronico, M. P. The effects of different reinforcing cues upon the free operant response rates of mental retardates and normal children of equated CA or MA. Unpublished doctoral dissertation, Rutgers University, 1963.

Bereiter, C. Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: Univer. Wisconsin Press, 1963. Pp. 3–20.

Bijou, S. W., & Baer, D. M. Some methodological contributions from a functional analysis of child development. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior.* Vol. 1. New York: Academic Press, 1963. Pp. 197–231.

Cairns, R. B. Antecedents of social reinforcer effectiveness. Progress Report No. M-4373, 1962, United States Public Health Service.

Craig, Grace J., & Myers, J. L. A developmental study of sequential two-choice decision making. *Child Development,* 1963, **34**, 483–493.

DeNike, L. D. The temporal relationship between awareness and performance in verbal conditioning. *Journal of Experimental Psychology,* 1964, **68**, 521–529.

Gewirtz, J. L., & Baer, D. M. Deprivation and satiation of social reinforcers as drive conditions. *Journal of Abnormal and Social Psychology,* 1958, **57**, 165–172. (a)

Gewirtz, J. L., & Baer, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *Journal of Abnormal and Social Psychology,* 1958, **56**, 49–56. (b)

Green, C., & Zigler, E. Social deprivation and the performance of retarded and normal children on a satiation type task. *Child Development,* 1962, **33**, 499–508.

Grossman, B. Parental warmth, child dependency and responsiveness to social reinforcement. Unpublished doctoral dissertation, Duke University, 1963.

Guilford, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.

Hartup, W. W. Friendship status and the effectiveness of peers as reinforcing agents. *Journal of Experimental Child Psychology,* 1964, **1**, 154–162.

Kelly, R. Experimenter characteristics and the effectiveness of social reinforcements of children's behavior. Unpublished doctoral dissertation, Purdue University, 1962.

Kelly, R., & Stephens, M. W. Comparison of different patterns of social reinforcement in children's operant learning. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 294–296.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

Lockwood, D. H. A developmental study of the efficacy of four cues as reinforcers in pre-school children. Unpublished doctoral dissertation, Rutgers University, 1963.

Long, E. R., Hammack, J. T., May, F., & Campbell, B. J. Intermittent reinforcement of operant behavior in children. *Journal of the Experimental Analysis of Behavior,* 1958, **1**, 315–339.

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: Univer. Wisconsin Press, 1963. Pp. 21–38.

McCullers, J. C., & Stevenson, H. W. Effects of verbal reinforcement in a probability learning situation. *Psychological Reports,* 1960, **7**, 439–445.

Parton, D. A., & Ross, A. O. Possible artifact in measurement of social reinforcement. *Psychological Reports,* 1964, **14**, 348.

Patterson, G. R., & Anderson, D. Peers as social reinforcers. *Child Development,* 1964, **35**, 951–960.

Patterson, G. R., & Hinsey, W. C. Investigations of some assumptions and characteristics of a procedure for instrumental conditioning in children. *Journal of Experimental Child Psychology,* 1964, **1**, 111–122.

Patterson, G. R., Littman, R. A., & Hinsey, W. C. Parental effectiveness as reinforcers in the laboratory and its relation to child rearing practices and child adjustment in the classroom. *Journal of Personality,* 1964, **32**, 180–199.

Shepps, R., & Zigler, E. Social deprivation and rigidity in the performance of organic and familial retardates. *American Journal of Mental Deficiency,* 1962, **67**, 262–268.

Spielberger, C. D. Theoretical and epistemological issues in verbal conditioning. In S. Rosenberg (Ed.), *Directions in psycholinguistics.* New York: Macmillan, 1965.

Stevenson, H. W. Social reinforcement with children as a function of CA, sex of *E* and sex of *S. Journal of Abnormal and Social Psychology,* 1961, **63**, 147–154.

Stevenson, H. W., & Cruse, D. B. The effectiveness of social reinforcement with normal and feebleminded children. *Journal of Personality,* 1961, **29**, 124–135.

Stevenson, H. W., & Fahel, Leila S. The effect of social reinforcement on the performance of institutionalized and noninstitutionalized normal and feebleminded children. *Journal of Personality,* 1961, **29**, 136–147.

Stevenson, H. W., Keen, Rachel, & Knights, R. M. Parents and strangers as reinforcing agents for children's performance. *Journal of Abnormal and Social Psychology,* 1963, **67**, 183–186.

Stevenson, H. W., & Knights, R. M. The effectiveness of social reinforcement after brief and extended institutionalization. *American Journal of Mental Deficiency,* 1962, **66**, 589–594. (a)

Stevenson, H. W., & Knights, R. M. Social reinforcement with normal and retarded children as a function of pretraining, sex of *E*, and sex of *S*. *American Journal of Mental Deficiency,* 1962, **66**, 866–871. (b)

Stevenson, H. W., & Odom, R. D. The effectiveness of social reinforcement following two conditions of social deprivation. *Journal of Abnormal and Social Psychology,* 1962, **65**, 429–431.

Stevenson, H. W., & Snyder, Leila C. Performance as a function of the interaction of incentive conditions. *Journal of Personality,* 1960, **28**, 1–11.

Walters, R. H., & Foote, Ann. A study of reinforcer effectiveness with children. *Merrill-Palmer Quarterly,* 1962, **8**, 149–157.

Walters, R. H., & Ray, E. Anxiety, social isolation, and reinforcer effectiveness. *Journal of Personality,* 1960, **28**, 358–367.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Zigler, E. Social deprivation and rigidity in the performance of feebleminded children. *Journal of Abnormal and Social Psychology,* 1961, **62**, 413–421.

Zigler, E. Rigidity and social reinforcement effects in the performance of institutionalized and noninstitutionalized normal and retarded children. *Journal of Personality,* 1963, **31**, 258–269.

Zigler, E. F., Hodgden, L., & Stevenson, M. W. The effect of support and nonsupport on the performance of normal and feebleminded children. *Journal of Personality,* 1958, **26**, 106–122.

Zigler, E., & Kanzer, P. The effectiveness of two classes of verbal reinforcers on the performance of middle-class and lower-class children. *Journal of Personality,* 1962, **30**, 157–163.

Zigler, E., & Williams, Joanne. Institutionalization and the effectiveness of social reinforcement. *Journal of Abnormal and Social Psychology,* 1963, **66**, 197–205.

(Received November 25, 1964)

Psychological Bulletin
1965, Vol. 64, No. 1, 74–79

$N = 1$

WILLIAM F. DUKES
University of California, Davis

Studies focused on the behavior of only one individual are examined. Despite the limitation on generalizing inherent in such studies, selected examples attest to their importance in the history of psychology. Their frequency in modern psychology is noted, together with their distribution across major topics. The usefulness of an N of 1 in research is viewed as extending beyond the single-case studies of clinician and personologists. An N of 1 is seen as also appropriate when, for the function considered, intersubject variability is low, when opportunities for observing a given class of events are limited, and when a supposed universal relationship is questioned and the obtained evidence is negative.

In the search for principles which govern behavior, psychologists generally confine their empirical observations to a relatively small sample of a defined population, using probability theory to help assess the generality of the findings obtained. Because this inductive process commonly entails some knowledge of individual differences in the behavior involved, studies employing only one subject ($N = 1$) seem somewhat anomalous. With no information about intersubject variability in performance, the general applicability of findings is indeterminate.

Although generalizations about behavior rest equally upon adequate sampling of both subjects and situations, questions about sampling most often refer to subjects. Accordingly, the term "$N = 1$" is used throughout the present discussion to designate the *reductio ad absurdum* in the sampling of subjects. It might, however, equally well (perhaps better, in terms of frequency of occurrence) refer to the limiting case in the sampling of situations—for example, the use of one maze in an investigation of learning, or a simple tapping task in a study of motivation. With respect to the two samplings, Brunswik (1956), foremost champion of the representative design of experiments, speculated:

> In fact, proper sampling of situations and problems may in the end be more important than proper sampling of subjects, considering the fact that individuals are probably on the whole much more alike than are situations among one another [p. 39].

As a corollary, the term $N = 1$ might also be appropriately applied to the sampling of experimenters. Long recognized as a potential source of variance in interview data (e.g., Cantril, 1944; Katz, 1942), the investigator has recently been viewed as a variable which may also influence laboratory results (e.g., McGuigan, 1963; Rosenthal, 1963).

Except to note these other possible usages of the term $N = 1$, the present paper is not concerned with one-experimenter or one-situation treatments, but is devoted, as indicated previously, to single-subject studies.

Despite the limitation stated in the first paragraph, $N = 1$ studies cannot be dismissed as inconsequential. A brief scanning of general and historical accounts of psychology will dispel any doubts about their importance, revealing, as it does, many instances of pivotal research in which the observations were confined to the behavior of only one person or animal.

SELECTIVE HISTORICAL REVIEW

Foremost among $N = 1$ studies is Ebbinghaus' (1885) investigation of memory. Called by some authorities "a landmark in the history of psychology . . . a model which will repay careful study [McGeoch & Irion, 1952, p. 1]," considered by others "a remedy . . . at least as bad as the disease [Bartlett 1932, p. 3]," Ebbinghaus' work establishe the pattern for much of the research o verbal learning during the past 80 years. Hi principal findings, gleaned from many self administered learning situations consisting o some 2,000 lists of nonsense syllables an 42 stanzas of poetry, are still valid sourc material for the student of memory. In a other well-known pioneering study of learning, Bryan and Harter's (1899) report on plateaus, certain crucial data were obtained from only one subject. Their letter-word-phrase analysis of learning to receive code was based on the record of only one student. Their notion of habit hierarchies derived in part from this analysis is, nevertheless, still useful in explaining why plateaus may occur.

Familiar even to beginning students of perception is Stratton's (1897) account of the confusion from and the adjustment to wearing inverting lenses. In this experiment according to Boring (1942), Stratton, with only himself as subject,

> settled both Kepler's problem of erect vision with an inverted image, and Lotze's problem of the role of experience in space perception, by showing that the "absolute" localization of retinal positions—up-down and right-left—are learned and consist of bodily orientation as context to the place of visual excitation [p. 237].

The role of experience was also under scrutiny in the Kelloggs' (1933) project of raising one young chimpanzee, Gua, in their home. (Although observations of their son's behavior were also included in their report, the study is essentially of the $N = 1$ type, since the "experimental group" consisted of one.) This attempt to determine whether early experience may modify behavior traditionally regarded as instinctive was for years a standard reference in discussions of the learning-maturation question.

Focal in the area of motivation is the balloon-swallowing experiment of physiologists Cannon and Washburn (1912) in which kymographic recordings of Washburn's stomach contractions were shown to coincide with his introspective reports of hunger pangs. Their findings were widely incorporated into psychology textbooks as providing an explanation of hunger. Even though in recent years greater importance has been attached to central factors in hunger, Cannon and Washburn's work continues to occupy a prominent place in textbook accounts of food-seeking behavior.

In the literature on emotion, Watson and Rayner's study (1920) of Albert's being conditioned to fear a white rat has been hailed as "one of the most influential papers in the history of American psychology" (Miller, 1960, p. 690). Their experiment, Murphy (1949) observes,

> immediately had a profound effect on American psychology; for it appeared to support the whole conception that not only simple motor habits, but important, enduring traits of personality, such as emotional tendencies, may in fact be 'built into' the child by conditioning [p. 261].

Actually the Albert experiment was unfinished because he moved away from the laboratory area before the question of fear removal could be explored. But Jones (1924) provided the natural sequel in Peter, a child who, through a process of active reconditioning, overcame a nonlaboratory-produced fear of white furry objects.

In abnormal psychology few cases have attracted as much attention as Prince's (1905) Miss Beauchamp, for years the model case in accounts of multiple personality. An excerpt from the Beauchamp case was recently included, along with selections from Wundt, James, Pavlov, Watson, and others, in a volume of 36 classics in psychology (Shipley, 1961). Perhaps less familiar to the general student but more significant in the history of psychology is Breuer's case (Breuer & Freud, 1895) of Anna O., the analysis of which is credited with containing "the kernel of a new system of treatment, and indeed a new system of psychology [Murphy, 1949, p. 307]." In the process of examining Anna's hysterical symptoms, the occasions for their appearance, and their origin, Breuer claimed that with the aid of hypnosis these symptoms were "talked away." Breuer's young colleague was Sigmund Freud (1910), who later publicly declared the importance of this case in the genesis of psychoanalysis.

There are other instances, maybe not so spectacular as the preceding, of influential $N = 1$ studies—for example, Yerkes' (1927) exploration of the gorilla Congo's mental activities; Jacobson's (1931) study of neuromuscular activity and thinking in an amputee; Culler and Mettler's (1934) demonstration of simple conditioning in a decorticate dog; and Burtt's (1932) striking illustration of his son's residual memory of early childhood.

TABLE 1

Total Distribution of $N = 1$ Studies (1939-1963)

Category	f	Examples
Maturation, development	29	Sequential development of prehension in a macaque (Jensen, 1961); smiling in a human infant (Salzen, 1963)
Motivation	7	Differential reinforcement effects of true, esophagal, and sham feeding in a dog (Hull, Livingston, Rouse, & Barker, 1951)
Emotion	12	Anxiety levels associated with bombing (Glavis, 1946)
Perception, sensory processes	25	Congenital insensitivity to pain in a 19-year-old girl (Cohen, Kipnis, Kunkle, & Kubzansky, 1955); figural aftereffects with a stabilized retinal image (Krauskopf, 1960)
Learning	27	Delayed recall after 50 years (Smith, 1963); imitation in a chimpanzee (Hayes & Hayes, 1952)
Thinking, language	15	"Idealess" behavior in a chimpanzee (Razran, 1961); opposite speech in a schizophrenic patient (Laffal & Ameen, 1959)
Intelligence	14	Well-adjusted congenital hydrocephalic with IQ of 113 (Teska, 1947); intelligence after lobectomy in an epileptic (Hebb, 1939)
Personality	51	Keats' personality from his poetry (McCurdy, 1944); comparison in an adult of P and R techniques (Cattell & Cross, 1952)
Mental health, psychotherapy	66	Multiple personality (Thigpen & Cleckley, 1954); massed practice as therapy for patient with tics (Yates, 1958)
Total	246	

Further documentation of the significant role of $N = 1$ research in psychological history seems unnecessary. A few studies, each in impact like the single pebble which starts an avalanche, have been the impetus for major developments in research and theory. Others, more like missing pieces from nearly finished jigsaw puzzles, have provided timely data on various controversies.

This historical recounting of "successful" cases is, of course, not an exhortation for restricted subject samplings, nor does it imply that their greatness is independent of subsequent related work.

Frequency and Range of Topics

In spite of the dated character of the citations—the latest being 1934—$N = 1$ studies cannot be declared the product of an era unsophisticated in sampling statistics, too infrequent in recent psychology to merit attention. During the past 25 years (1939–1963) a total of 246 $N = 1$ studies, 35 of them in the last 5-year period, have appeared in the following psychological periodicals: the *American Journal of Psychology*, *Journal of Genetic Psychology*, *Journal of Abnormal and Social Psychology*, *Journal of Educational Psychology*, *Journal of Comparative and Physiological Psychology*, *Journal of Experimental Psychology*, *Journal of Applied Psychology*, *Journal of General Psychology*, *Journal of Social Psychology*, *Journal of Personality*, and *Journal of Psychology*. These are the journals, used by Bruner and Allport (1940) in their survey of 50 years of change in American psychology, selected as significant for and devoted to the advancement of psychology as science. (Also used in their survey were the *Psychological Review*, *Psychological Bulletin*, and *Psychometrika*, excluded here because they do not ordinarily publish original empirical work.) Although these 246 studies constitute only a small per cent of the 1939–1963 journal articles, the absolute number is noteworthy and is sizable enough to discount any notion that $N = 1$ studies are a phenomenon of the past.

When, furthermore, these are distributed, as in Table 1, according to subject matter, they are seen to coextend fairly well with the range of topics in general psychology. As might be expected, a large proportion of them fall into the clinical and personality areas. One cannot, however, explain away $N = 1$ studies as case histories contributed by clinicians and personologists occupied less with establishing generalizations than with exploring the uniqueness of an individual and understanding his total personality. Only about 30% (74) are primarily oriented toward the individual, a figure which includes not only works in the "understanding" tradition, but also those treating the individual as a universe of responses and applying traditionally nomothetic techniques to describe and predict individual behavior (e.g., Cattell & Cross, 1952; Yates, 1958).

In actual practice, of course, the two orientations—toward uniqueness or generality—are more a matter of degree than of mutual exclusion, with the result that in the literature surveyed purely idiographic research is extremely rare. Representative of that approach are Evans' (1950) novel-like account of Miller who "spontaneously" recovered his sight after more than 2 years of blindness, Rosen's (1949) "George X: A self-analysis by an avowed fascist," and McCurdy's (1944) profile of Keats.

Rationale for $N = 1$

The appropriateness of restricting an idiographic study to one individual is obvious from the meaning of the term. If uniqueness is involved, a sample of one exhausts the population. At the other extreme, an N of 1 is also appropriate if complete population generality exists (or can reasonably be assumed to exist). That is, when between-individual variability for the function under scrutiny is known to be negligible or the data from the single subject have a point-for-point congruence with those obtained from dependable collateral sources, results from a second subject may be considered redundant. Some $N = 1$ studies may be regarded as approximations of this ideal case, as for example, Heinemann's (1961) photographic measurement of retinal images and Bartley and Seibel's (1954) study of entoptic stray light, using the flicker method.

A variant on this typicality theme occurs when the researcher, in order to preserve some kind of functional unity and perhaps to dramatize a point, reports in depth one case which exemplifies many. Thus Eisen's (1962) description of the effects of early sensory deprivation is an account of one quondam hard-of-hearing child, and Bettelheim's (1949) paper on rehabilitation a chronicle of one seriously delinquent child.

In other studies an N of 1 is adequate because of the dissonant character of the findings. In contrast to its limited usefulness in *establishing* generalizations from "positive" evidence, an N of 1 when the evidence is "negative," is as useful as an N of 1,000 in *rejecting* an asserted or assumed universal relationship. Thus Krauskopf's (1960) demonstration with one stopped-image subject eliminates motion of the retinal image as necessary for figural aftereffects; and Lenneberg's (1962) case of an 8-year-old boy who lacked the motor skills necessary for speaking but who could understand language makes it "clear that hearing oneself babble is not a necessary factor in the acquisition of understanding . . . [p. 422]." Similarly Teska's (1947) case of a congenital hydrocephalic, 6½ years old, with an IQ of 113, is sufficient evidence to discount the notion that prolonged congenital hydrocephaly results in some degree of feeblemindedness.

While scientists are in the long run more likely to be interested in knowing *what is* than *what is not* and more concerned with how many exist or in what proportion they exist than with the fact that at least one exists, one negative case can make it necessary to revise a traditionally accepted hypothesis.

Still other $N = 1$ investigations simply reflect a limited opportunity to observe. When the search for lawfulness is extended to infrequent "nonlaboratory" behavior, individuals in the population under study may be so sparsely distributed spatially or temporally that the psychologist can observe only one case, a report of which may be useful as a part of a cumulative record. Examples of this include cases of multiple personality (Thigpen & Cleckly, 1954), unilateral color blindness (Graham, Sperling, Hsia, & Coulson, 1961) congenital insensitivity to pain (Cohen et al., 1955), and mental deterioration following carbon monoxide poisoning (Jensen, 1950). Situational complexity as well as subject sparsity may limit the opportunity to observe. When the situation is greatly extended in time, requires expensive or specialized training for the subject, or entails intricate and difficult to administer controls, the investigator may, aware of their exploratory character, restrict his observations to one subject. Projects involving home-raising a chimpanzee (Hayes & Hayes, 1952) or testing after 16 years for retention of material presented during infancy (Burtt, 1941) would seem to illustrate this use of an N of 1.

Not all $N = 1$ studies can be conveniently fitted into this rubric; nor is this necessary. Instead of being oriented either toward the person (uniqueness) or toward a global theory (universality), researchers may sometimes simply focus on a problem. Problem-centered research on only one subject may, by clarifying questions, defining variables, and indicating approaches, make substantial contributions to the study of behavior. Besides answering a specific question, it may (Ebbinghaus' work, 1885, being a classic example) provide important groundwork for the theorists.

Regardless of rationale and despite obvious limitations, the usefulness of $N = 1$ studies in psychological research seems, from the preceding historical and methodological considerations, to be fairly well established. (See Shapiro, 1961, for an affirmation of the value of single-case investigations in fundamental clinical psychological research.) Finally, their status in research is further secured by the statistician's assertion (McNemar, 1940) that:

> The statistician who fails to see that important generalizations from research on a single case can ever be acceptable is on a par with the experimentalist who fails to appreciate the fact that some problems can never be solved without resort to numbers [p. 361].

REFERENCES

Bartlett, F. C. *Remembering.* Cambridge, England: University Press, 1932.

Bartley, S. H., & Seibel, Jean L. A further study of entoptic stray light. *Journal of Psychology,* 1954, **38**, 313–319.

Bettelheim, B. H. A study in rehabilitation. *Journal of Abnormal and Social Psychology,* 1949, **44**, 231–265.

Boring, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century, 1942.

Breuer, J., & Freud, S. Case histories. (Orig. publ. 1895; trans. by J. Strachey) In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud.* Vol. 2. London: Hogarth Press, 1955. Pp. 19–181.

Bruner, J. S., & Allport, G. W. Fifty years of change in American psychology. *Psychological Bulletin,* 1940, **37**, 757–776.

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

Bryan, W. L., & Harter, N. Studies on the telegraphic language. The acquisition of a hierarchy of habits. *Psychological Review,* 1899, **6**, 345–375.

Burtt, H. E. An experimental study of early childhood memory. *Journal of Genetic Psychology,* 1932, **40**, 287–295.

Burtt, H. E. An experimental study of early childhood memory: Final report. *Journal of Genetic Psychology,* 1941, **58**, 435–439.

Cannon, W. B., & Washburn, A. L. An explanation of hunger. *American Journal of Physiology,* 1912, **29**, 441–454.

Cantril, H. *Gauging public opinion.* Princeton: Princeton Univer. Press, 1944.

Cattell, R. B., & Cross, K. P. Comparison of the ergic and self-sentiment structures found in dynamic traits by R- and P- techniques. *Journal of Personality,* 1952, **21**, 250–271.

Cohen, L. D., Kipnis, D., Kunkle, E. C., & Kubzansky, P. E. Observations of a person with congenital insensitivity to pain. *Journal of Abnormal and Social Psychology,* 1955, **51**, 333–338.

Culler, E., & Mettler, F. A. Conditioned behavior in a decorticate dog. *Journal of Comparative Psychology,* 1934, **18**, 291–303.

Ebbinghaus, H. *Über das Gedächtnis.* Leipzig: Duncker & Humblot, 1885.

Eisen, N. H. Some effects of early sensory deprivation on later behavior: The quondam hard-of-hearing child. *Journal of Abnormal and Social Psychology,* 1962, **65**, 338–342.

Evans, Jean. Miller. *Journal of Abnormal and Social Psychology,* 1950, **45**, 359–379.

Freud, S. The origin and development of psychoanalysis. *American Journal of Psychology,* 1910, **21**, 181–218.

Glavis, L. R., Jr. Bombing mission number fifteen. *Journal of Abnormal and Social Psychology,* 1946, **41**, 189–198.

Graham, C. H., Sperling, H. G., Hsia, Y., & Coulson, A. H. The determination of some visual functions of a unilaterally color-blind subject: *Journal of Psychology,* 1961, **51**, 3–32.

Hayes, K. J., & Hayes, Catherine. Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology,* 1952, **45**, 450–459.

Hebb, D. O. Intelligence in man after large removals of cerebral tissue: Defects following right temporal lobectomy. *Journal of General Psychology,* 1939, **21**, 437–446.

HEINEMANN, E. G. Photographic measurement of the retinal image. *American Journal of Psychology,* 1961, **74**, 440–445.

HULL, C. L., LIVINGSTON, J. R., ROUSE, R. O., & BARKER, A. N. True, sham, and esophageal feeding as reinforcements. *Journal of Comparative and Physiological Psychology,* 1951, **44**, 236–245.

JACOBSON, E. Electrical measurements of neuromuscular states during mental activities: VI. A note on mental activities concerning an amputated limb. *American Journal of Physiology,* 1931, **96**, 122–125.

JENSEN, G. D. The development of prehension in a macaque. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 11–12.

JENSEN, M. B. Mental deterioration following carbon monoxide poisoning. *Journal of Abnormal and Social Psychology,* 1950, **45**, 146–153.

JONES, MARY C. A laboratory study of fear: The case of Peter. *Journal of Genetic Psychology,* 1924, **31**, 308–315.

KATZ, D. Do interviewers bias poll results? *Public Opinion Quarterly,* 1942, **6**, 248–268.

KELLOGG, W. N., & KELLOGG, LUELLA. *The ape and the child.* New York: McGraw-Hill, 1933.

KRAUSKOPF, J. Figural after-effects with a stabilized retinal image. *American Journal of Psychology,* 1960, **73**, 294–297.

LAFFAL, J., & AMEEN, L. Hypotheses of opposite speech. *Journal of Abnormal and Social Psychology,* 1959, **58**, 267–269.

LENNEBERG, E. H. Understanding language without ability to speak: A case report. *Journal of Abnormal and Social Psychology,* 1962, **65**, 419–425.

MCCURDY, H. G. *La belle dame sans merci. Character and Personality,* 1944, **13**, 166–177.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

MCGUIGAN, F. J. The experimenter: A neglected stimulus object. *Psychological Bulletin,* 1963, **60**, 421–428.

MCNEMAR, Q. Sampling in psychological research. *Psychological Bulletin,* 1940, **37**, 331–365.

MILLER, D. R. Motivation and affect. In Paul H. Mussen (Ed.), *Handbook of research methods in child development.* New York: Wiley, 1960. Pp. 688–769.

MURPHY, G. *Historical introduction to modern psychology.* New York: Harcourt, Brace, 1949.

PRINCE, M. *The dissociation of a personality.* New York: Longmans, Green, 1905.

RAZRAN, G. Raphael's "idealess" behavior. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 366–367.

ROSEN, E. George X: The self-analysis of an avowed fascist. *Journal of Abnormal and Social Psychology,* 1949, **44**, 528–540.

ROSENTHAL, R. Experimenter attributes as determinants of subjects' responses. *Journal of Projective Techniques,* 1963, **27**, 324–331.

SALZEN, E. A. Visual stimuli eliciting the smiling response in the human infant. *Journal of Genetic Psychology,* 1963, **102**, 51–54.

SHAPIRO, M. B. The single case in fundamental clinical psychological research. *British Journal of Medical Psychology,* 1961, **34**, 255–262.

SHIPLEY, T. (Ed.) *Classics in psychology.* New York: Philosophical Library, 1961.

SMITH, M. E. Delayed recall of previously memorized material after fifty years. *Journal of Genetic Psychology,* 1963, **102**, 3–4.

STRATTON, G. M. Vision without inversion of the retinal image. *Psychological Review,* 1897, **4**, 341–360, 463–481.

TESKA, P. T. The mentality of hydrocephalics and a description of an interesting case. *Journal of Psychology,* 1947, **23**, 197–203.

THIGPEN, C. H., & CLECKLEY, H. A case of multiple personality. *Journal of Abnormal and Social Psychology,* 1954, **49**, 135–151.

WATSON, J. B., & RAYNER, ROSALIE. Conditioned emotional reactions. *Journal of Experimental Psychology,* 1920, **3**, 1–14.

YATES, A. J. The application of learning theory to the treatment of tics. *Journal of Abnormal and Social Psychology,* 1958, **56**, 175–182.

YERKES, R. M. The mind of a gorilla. *Genetic Psychology Monographs,* 1927, **2**, 1–193.

(Received December 3, 1964)

Psychological Bulletin
1965, Vol. 64, No. 1, 80

COMMENTS ON "NEED STATES AND ACTIVITY LEVEL":

A CRITIQUE

L. GLENN COLLINS

Mental Health Research Institute, Fort Steilacoom, Washington

A note on the paper by Baumeister, Hawkins, and Cromwell (1964) to point out the risk inherent in relying on secondary sources. These reviewers stated that copulation by the male rat does not result in decreased activity levels and cited Morgan and Stellar (1950) as support for this statement. The relevant paragraph in Morgan and Stellar refers to an experiment by Stone and Barker (1934) which does not support the reviewers' statement.

An important point in Baumeister, Hawkins, and Cromwell's recent paper (1964) is that exponents of reinforcement theories find it difficult to explain the finding that copulation does not serve to reduce the "running wheel" activity of the male rat. They gave as a reference Morgan and Stellar (1950). Lacking a page reference but being familiar with that book, this reviewer was able to find the section to which (it was assumed) the reviewers were referring.

> The male rat does not show the periodicity characteristic of the activity of the female rat but maintains, rather, a relatively constant level of activity. There is, moreover, little or no relation between the general activity of a male rat and the strength of its sexual drive as measured by the readiness with which it will avail itself of a receptive female (Stone and Barker). In this respect, therefore, there is no important relation between activity and strength of drive.
>
> Of interest also is the fact that reduction in sexual drive has no visible effect on acivity, for animals allowed a period with a female rat in "heat" are just as active in a revolving drum as before [Morgan & Stellar, 1950, p. 371].

Referring to Stone and Barker's paper (1934), nothing was found in the text or tables to support the statement by Morgan and Stellar that copulation does not serve to reduce the activity of the male. This statement may or may not be valid but the reference cited fails to support it. This seems to be another instance in which reviewers have unfortunately depended upon an unreliable secondary source and unwittingly perpetrated a misinterpretation.

REFERENCES

Baumeister, A., Hawkins, W., & Cromwell, R. Need states and activity level. *Psychological Bulletin*, 1964, **61**, 438–453.

Morgan, C. T., & Stellar, E. *Physiological psychology*. New York: McGraw-Hill, 1950.

Stone, C. P., & Barker, R. G. Spontaneous activity, direct and indirect measures of sexual drive in adult male rats. *Proceedings of the Society for Experimental Biology and Medicine*, 1934, **32**, 195–199.

(Received March 12, 1965)

VOL. 64, NO. 2 AUGUST 1965

Psychological Bulletin

TOWARD THE EXPERIMENTAL CONTROL OF DREAMING:

A REVIEW OF THE LITERATURE

CHARLES T. TART [1]

Laboratory of Hypnosis Research, Stanford University

The study of nocturnal dreaming has been mainly a matter of correlational studies. This article brings together and reviews those scattered studies in which an attempt was made to actively influence dreaming. Despite a great diversity of methods, aims, and results, it is apparent that dreaming can be experimentally manipulated, and thus it may be possible to develop a functional experimental approach in which dreams become more immediate and observable aspects of behavior, rather than only memories.

Scientific research is carried out with two basic kinds of procedures. One is the observational or correlational approach, where we observe "what happens naturally" and try to make sense of it by finding relations or correlations between various aspects of our observations. This approach leads to statements of the order, "subjects classified as Sensitizers report recalling dreaming significantly more frequently than those classified as Repressors" on the MMPI (Tart, 1962). The other approach is the functional method, where we actively manipulate one factor or variable and observe the effect of some dependent variable. This approach leads to statements of the order, "A large dose of alcohol reduces the amount of EEG stage 1-REM time and inferred dreaming in normal Subjects . . ." (Gresham, Webb, & Williams, 1963). Both approaches are used in most fields of science, as they are complementary.

Research into the nature of dreaming has been, by and large, a matter of correlational research (ignoring for the moment the last decade's research using the electroencephalogram, EEG, and rapid eye movement, REM, techniques). Researchers asked subjects how often they dreamed in color, whether their dreams were mainly pleasant or unpleasant, and the like, and then attempted to correlate these observed characteristics with personality characteristics, intelligence, sex, age, etc. Or the patient brought in a spontaneous dream, and the therapist-researcher then attempted to "interpret" it in the light of what he knew about the patient.

We have, of course, learned a great deal about the nature of dreaming as a result of these correlational studies, and undoubtedly we will continue to learn from them. When we can complement correlational studies with functional studies, however, we shall advance our understanding of dreaming much more rapidly.

The purpose of the present paper is to review and bring together for the first time those studies in which an experimental attempt to influence either the content or the process of nocturnal dreaming was made, in order to assess what possibilities they suggest for functional research on dreaming.

The past decade has seen an enormous increase in our knowledge of sleep and dreaming resulting from application of EEG measures, and a brief summary of this body of results will be presented in order to provide an up-to-date background for assessing the studies to be reviewed herein. More detailed reviews and evaluations of the more than 75 studies contributing to the past decade's new knowledge of dreaming have been published elsewhere (Kamiya, 1961; Kleitman, 1960, 1963; Oswald, 1962, 1964; Snyder, 1963).

---

[1] This investigation was supported (in part) by a Public Health Service postdoctoral fellowship (1-F2-MH-14,622-01) from the National Institute of Mental Health. I also wish to thank Leslie M. Cooper and Ernest R. Hilgard for their helpful comments during the preparation of this manuscript.

Sleep will be defined in this paper as a state of the organism indicated (in human subjects) by one of the four EEG stages (Dement & Kleitman, 1957a, 1957b). The Stage 1 pattern consists of an irregular mixture of theta waves (4–8 cps), alphoid activity (waves of 1 to 2 cps slower than the subject's waking alpha), and occasional alpha waves (8–13 cps). Stage 2 contains spindle activity (14 cps) in addition to the above, and Stages 3 and 4 contain increasingly larger proportion (up to 100%) of delta waves (1–3 cps, high amplitude) in addition to spindle activity. The exact division between Stages 2, 3, and 4 are arbitrary, based on the percentage of delta waves in given epochs. The Stage 1 pattern is readily distinguishable from the other stages by its total lack of spindles and delta waves.

Stages 1 through 4 were initially conceived of as comprising a continuum from "light" to "deep" sleep (Aserinsky & Kleitman, 1953, 1955; Dement, 1955), but as other measures of the "depth" of sleep contradict this conception (Berger, 1961; Hawkins, Puryear, Wallace, Deal, & Thomas, 1962; Kamiya, 1961; Snyder, 1963; Williams, Tepas, & Morlock, 1962), this paper will treat sleep as being of two qualitatively distinct types, namely Stage 1 as one type and Stages 2, 3, and 4 as the other type. Distinctions between Stages 2, 3, and 4 will not be made in this paper, and they will be collectively referred to as Nonstage 1 sleep.

If subjects are awakened from the two types of sleep and asked to report what they have been experiencing, the reports may be classified into two rather distinct types. One type, awakenings from Stage 1 sleep or shortly (within, roughly, 10–15 minutes) after Stage 1 sleep has changed to Nonstage 1 sleep, possesses the characteristics traditionally associated with the experience of dreaming (Foulkes, 1962; Rechtschaffen, Verdone, & Wheaton, 1963). Reports from Nonstage 1 sleep seem more like "thinking," and are generally called thinking by the subjects—these same subjects generally refer to their Stage 1 experiences as dreams. The psychological differences reported so far are quantitative, rather than being completely dichotomous, but generally give the impression of distinct types of experiences.

Stage 1 sleep is almost always accompanied by binocularly synchronous rapid eye movements (REMs), and the evidence is very convincing that these are closely associated with the content of the dream, if not actual scanning movements of the dream imagery (Berger & Oswald, 1962a; Dement & Wolpert, 1958; Roffwarg, Dement, Muzio, & Fisher, 1962). Such REMs have not been reported in Nonstage 1 sleep, although there are some slow, rolling movements (Kamiya, 1961).

In view of these findings, the following theoretical position will be taken in this paper as an aid in surveying the literature on functional studies of dreaming: that an experientially distinct type of phenomenon occurs concurrently with the presence of Stage 1 sleep, which phenomenon will be called Stage 1 dreaming, or just dreaming. The mental phenomena of Nonstage 1 sleep will not be considered in this paper. Further, it is assumed that the experience of Stage 1 dreaming is essentially continuous [2] during the presence of Stage 1 EEG, whether the subject can always recall this experience on waking or not. This position is, in the author's opinion, supported by all the studies using the EEG and REM technique, and directly refuted by none.

The term "process" of Stage 1 dreaming is used to indicate the ongoing psychophysiological activity which is observed on a physiological level as the Stage 1 EEG pattern and on an experimental level as the experience of dreaming (observed directly by the subject and indirectly via the dream report). The terms "dream content," "reported dream," or simply "dream" refer to the particulars of the experience.

---

[2] Within a continuous period of Stage 1 EEG, the content of the experienced dream may be divided into several distinct episodes so that, in a sense, there are several distinct "dreams" within a continuous period of dreaming. Dement and Wolpert (1958) present some evidence that such changes of topic may be accompanied by a gross body movement on the part of the subject.

For normal subjects, Stage 1 dreaming and Nonstage 1 sleep alternate in a regular cyclic fashion referred to as the sleep-dream cycle. As a subject falls asleep there is generally a brief (a few seconds to a minute or two) period of Stage 1, without REMs, but subjects' reports indicate that this is apparently a period of hypnagogic imagery, rather than typical dreaming (Dement & Kleitman, 1957a; Oswald, 1962). At approximately 90-minute intervals through the night there are periods of Stage 1 dreaming, each dream period generally being longer than the preceding one. The first Stage 1 period may last for 10 minutes; the fourth or fifth one may last as long as 50 minutes. Altogether, Stage 1 dreaming occupies between 20 and 30% of the total sleep time of most young adults, spread over three to six Stage 1 periods. While the exact percentage of dream time and the number of cycles varies from subject to subject, for a given subject the sleep-dream cycle is generally quite stable from night to night (Dement, 1960; Dement & Kleitman, 1957b; Kleitman, 1960; Wood, 1962).

With this picture in mind of Stage 1 dreaming occurring regularly through every night of sleep, let us now review those studies which have attempted to affect dreaming. Many of these studies were carried out before the discovery of the sleep-dream cycle and so lack precision, but do offer some data.

The studies will be reviewed under three convenient, but not mutually exclusive headings, namely: (*a*) variables affecting the content of dreaming, (*b*) variables affecting the process of dreaming, and (*c*) discrimination during sleep.

Variables Affecting the Content of Dreaming

Variables affecting dreaming are of two types, namely, those applied before the onset of sleep, and those applied concurrently with sleep and dreaming. The latter type will be discussed first.

Naturally occurring stimuli, both external and internal (e.g., thunder, breezes, stomach cramps, falling out of bed), are sometimes incorporated into the content of dreams. There are many anecdotal accounts of this in the older literature (Ellis, 1911; Lhermitte, 1931; Schatzman, 1925; Spitta, 1882; Sully, 1905; Thompson, 1914; Walsh, 1920), some of which are cited by Freud (1954; i.e., works by Hennings, Hoffbauer, Jessen, and Meier). Such accounts tell us little about the effectiveness of stimuli in modifying dream content, however, as they provide no baseline of how many naturally occurring stimuli were not incorporated, nor do they provide any systematic data on the manner in which the incorporated stimuli were represented in the dream.

De Sanctis and Neyroz (1902), Max (1935), and Vold (1912) experimentally presented stimuli to sleeping subjects, as did Claviere, Hildebrandt, Maury, Stepanow, Straecke, Vaschide (cited by Ramsey, 1953), and Menuier (cited by Ellis, 1911). They also demonstrated that stimuli are sometimes incorporated into dreams. Despite great variation in method from study to study, there seems to be a consensus in these studies that tactual stimuli are more effective in affecting dream content than stimuli in other sensory modalities.

Both the anecdotal and early experimental studies were, in the light of what is now known about the sleep-dream cycle, beset with an inherent lack of precision, as data concerning Stage 1 sleep and Nonstage 1 sleep were indiscriminately mixed together. A recent study (Dement & Wolpert, 1958), using the EEG and REM technique, has greatly increased our knowledge here. Briefly, a pure tone was incorporated in 9% of the trials, a bright light in 23%, and drops of water in 42%. As the water is a tactual stimulus, this is in agreement with the older studies. When these stimuli were presented during Nonstage 1 sleep, Stage 1 dreaming was *not* initiated, contrary to the old idea that dreaming was *caused* by stimulation, and when the subjects were awakened a few minutes later they had no recollection of the stimuli (or of dreaming).

An important and methodologically sophisticated study of stimulus incorporation into Stage 1 dreams is that of Berger (1963). Spoken first names were presented at randomly chosen times during Stage 1 dreaming. Some of the names were of persons with whom the subjects had important, emotional relationships, while others were neutral in this respect. Incorporation was objectively judged by having both the subjects and an independent judge later try to match reported dream content with stimulus names. The stimulus names were matched with the correct dream reports with considerably better-than-chance accuracy by both the subjects and the independent judge. The emotionally meaningful and the neutral names were incorporated with equal frequency. Incorporation was not associated with detectable EEG changes, or GSR responses. Berger reported the incorporations as falling into several descriptive categories, in order of decreasing frequency, as follows: (*a*) assonance, where a dream element or word sounded like the stimulus name, a sort of "clang association"; (*b*) direct incorporation, where the dreamer reported hearing the stimulus name in the dream; (*c*) associational incorporation, where a relatively direct personal association to the stimulus name appeared in the dream; and (*d*) representational association, where the stimulus name was represented or symbolized by an object or action in the dream.

There have been some studies of the incorporation of stimuli into hypnotic dreams (Benussi, 1927; Klein, 1930), but as there is strong evidence that hypnotic dreams may be quite different from Stage 1 dreams (Barber, 1962; Tart, 1964b, 1965) they will not be reviewed here.

Some experimenters applied various stimuli to their subjects before the latter went to sleep (Cubberly, 1923; Fisher, 1954, 1960; Malamud & Lindner, 1931; Monroe, 1899; Poetzl, cited by Ramsey, 1953; Renshaw, Miller, & Marquis, 1933; Vold, 1912) and found that the stimuli sometimes appeared in the content of dreams reported the following day, often in "distorted" or "disguised" form. Such studies are difficult to evaluate, as there is a contaminating factor of waking suggestion to dream about the stimuli implicit in the design of stimulating subjects and then asking them to record their dreams. More serious difficulties in evaluating these studies arise from the heavy demand characteristics (Orne, 1959, 1962) undoubtedly operative as a result of the psychoanalytic orientation of many of the experimenters. About the most that can be concluded from these studies is that subjects will bring in "dreams" which seem to incorporate the stimulus material in disguised or distorted form, but this fact does not necessarily prove the validity of psychoanalytic theories of dream formation. These studies are also subject to other criticisms, discussed below.

A number of studies apparently demonstrated that posthypnotic suggestion could control the content of nocturnal dreaming to a high degree (Fisher, 1953; Nachmansohn, 1951; Newman, Katz, & Rubenstein, 1960; Schrötter, 1951), but as they did not use EEG monitoring of sleep, they were not controlled for the possibility that the dreams affected were hypnotic dreams interspersed with sleep, rather than ordinary nocturnal dreams. That such an event can happen, without the subject being aware of it, has been shown in other studies (Barber, 1962; Schiff, Bunney, & Freedman, 1961; Tart, 1964b). These studies also are ambiguous in that effects on Stage 1 dreaming and Nonstage 1 mental activity would be indiscriminately mixed, and much of the early Stage 1 dreaming of the night was undoubtedly forgotten by the time the subjects awoke in the morning.

Stoyva (1961), using the EEG and REM technique, was the first to demonstrate convincingly that the content of Stage 1 dreaming could be influenced and manipulated to a very high degree in some subjects by means of posthypnotic suggestion. The content of Nonstage 1 mental activity was also affected, although not as markedly. Tart (1964b) independently confirmed Stoyva's finding. As with Stoyva's subjects, the Stage 1 dreams of some subjects were not affected at all, while those of other subjects were almost totally controlled by the posthypnotically suggested content. Stoyva also found that explicit waking suggestion affected dream content, as had been reported earlier by Fisher (1953) and Titchener (1895), but not as strongly as posthypnotic suggestion.

Wood (1962) found that social isolation for a period of 1 day affected the content of Stage 1 dreaming occurring that night. The subjects seemed to be compensating for the isolation by increased "social intercourse" in their dreams.

The laboratory situation itself seems to have some effect on Stage 1 dreaming, especially when psychiatric patients are used as subjects (Whitman, Pierce, Maas, & Baldridge, 1962). Feelings about the experimenter and about being experimented on appeared in the reported dream content, often in "disguised" form. However, it has been reported (Dement, Kahn, & Roffwarg, in press) that this effect diminishes rapidly after a few nights in the laboratory, with normal subjects. About 24% of first-night dreams were concerned with the experimental situation directly, but this fell to 9% for later nights.

Several studies have been carried out to investigate the effects of various drugs on dream content, for example, chloropromazine and promazine (Lesse, 1959), meprobamate (Whitman, Pierce, & Maas, 1960), and imipramine and prochlorperazine (Whitman, Pierce, Maas, & Baldridge, 1961). While these studies indicate some general changes in content, this area is still in a very exploratory stage, and no definite conclusions can be drawn yet.

At present, then, posthypnotic suggestion seems to be the most powerful and precise method for affecting dream content, although its use is restricted to a minority of subjects. The use of external stimuli presented to the dreaming subject offers promise of strongly affecting dream content, but much research remains to be done in this area, using the EEG-REM technique, to discover the optimal types of stimuli and ways of presenting them.

Variables Affecting the Process of Dreaming

There are two powerful techniques which result in almost total disruption of the sleep-dream cycle. The first is total sleep deprivation, which has received considerable attention in recent years (Berger & Oswald, 1962b; Brauchi & West, 1959; Luby, Frohman, Grisell, Lenzo, & Gottlieb, 1960; Luby, Grisell, Frohman, Lees, Cohen, & Gottlieb, 1962; Mirsky & Cardon, 1962; Rodin, Luby, & Gottlieb, 1962; Simon, 1961; West, Janszen, Lester, Cornelisoon, 1962; Williams, Granda, Jones, Lubin, & Armington, 1962; Williams, Hammack, Daly, Dement, & Lubin, 1964; Williams, Morris, & Lubin, 1962). The second is selective deprivation of certain aspects of sleep, Dement's (1960) "dream-deprivation" technique where subjects are awakened whenever a Stage 1 pattern appears, and Stage 4 deprivation (Agnew, Webb, & Williams, 1964) carried out in the same manner.

Dement found that if, on several consecutive nights, a subject was awakened whenever Stage 1 dreaming began, the subject would have to be awakened more and more frequently to prevent Stage 1 dreaming as the nights went on. Furthermore, if the subject was allowed several undisturbed nights of sleep following this procedure, his Stage 1 dream time was significantly increased for several nights. Control awakenings scattered randomly through Nonstage 1 sleep produced no such effect. Dement concluded that there was a need for a given amount of Stage 1 dreaming each night, although whether this is a need for the experience of dreaming or the physiological state represented by Stage 1 EEG is unknown. Agnew et al. (1964) followed the same procedure for Stage 4 of sleep, although the stimuli given their subjects whenever Stage 4 sleep started often served to shift them to a different stage of sleep without awakening. They found results similar to Dement's, that is, more and more need to stimulate the subjects to get them out of Stage 4 as the deprivation nights went on, and an increase in Stage 4 sleep time on undisturbed nights following this procedure.

As both the sleep-deprivation and dream-deprivation procedures represent a total disruption of the sleep-dream cycle, rather than a selective effect on dreaming per se, they will not be further discussed here.

The laboratory situation itself has an effect on dream process, just as it does on dream content. On a subject's first night in the laboratory he is likely to miss his first Stage 1 period of the night. This phenomenon has been called the "first-night effect" (Dement, 1955; Dement et al., in press; Rechtschaffen & Verdone, 1964; Snyder, 1963), and it is now customary among dream researchers to have a subject's first night count only as an adaptation night, rather than taking any precise measurements on it.

Social isolation also affects the process of

Stage 1 dreaming. Wood (1962) found that a day of social isolation resulted in a 60% increase of Stage 1 dream time for the first two dream cycles of the experimental night, while the number of REMs/minute during this dreaming was decreased.

Gresham et al. (1963) studied the effect of alcohol and caffeine on Stage 1 dream time, and found that alcohol produced a significant decrease in Stage 1 dream time (a mean decrease of 22 minutes for the first 5 hours of sleep), while caffeine, in the dosage used, had no effect. Rechtschaffen and Maron (1964) found that a mixture of dextro amphetamine sulfate and pentobarbitol sodium may halve Stage 1 dream time in many subjects. Dement (cited in Snyder, 1963) found that sodium amytal may also halve dream time. Heptabaritone (Oswald, Berger, Jaramillo, Keddie, Olley, & Plunkett, 1963) also reduces Stage 1 dream time.

Tactual stimuli, in the form of the hardness of the surface on which the subject sleeps (a feather bed at one extreme and a board at the other) have been reported to affect sleep (Suckling, Koening, Hoffman, & Brooks, 1957), but the data are not presented in a form which allows careful assessment of the effect on Stage 1 dreaming.

Stoyva (1961) found that posthypnotic suggestions intended to influence Stage 1 dream *content* also affected the duration of Stage 1 dreaming. In those of his subjects whose dream content was generally influenced by the suggestions in every dream of the night, Stoyva found a statistically significant decrease in Stage 1 dream time for the early part of the night, what he termed the "shortening effect." For the first two Stage 1 periods of the night, this effect was an average 7-minute decrease, 21% with a maximum decrease of 14 minutes.

Offering subjects money if they can dream more or less on different nights also has some effect (Rechtschaffen & Verdone, 1964), although the difference between the two conditions was only an average of 6.6 minutes per night (±3% compared to base-line nights).

A particularly interesting problem is that of exerting control over waking from sleep. Ladd (1892), speaking of whatever "mechanism" caused sleep, reported that "I 'set' this mechanism so that it will dip down into sleep and dream-life, with a gradual curvature, as it were, and then come out of dream-life in an instant, i.e., by a steep curve." He provides no other information as to the nature of this sleep "mechanism" or how he went about "setting" it, although one might guess that he used some form of autosuggestion. Nor does he report how successful this technique was, although he mentions that it was more successful for arousing himself shortly after going to sleep than for late in sleep.

A very intriguing report on waking from sleep is that of Frobenius (1927), which shows that subjects can wake from sleep at a randomly preselected target time on most attempts. His subjects usually awoke within 10 minutes of the selected time, and felt rather disoriented. They had no idea as to how they had awakened, nor did the content of their dreams seem to reflect a waking process in any way on those occasions when they awoke from dreams. Elder (1941), in a brief abstract, reports that he had confirmed Frobenius' results, but provides no details. This material will soon be submitted for publication, however by J. Elder.[3] Ehrenwald (1923) also reports some anecdotal data on waking from sleep at a preset time.

Other investigators (Calkins, 1893; Dittborn, 1963; Nelson, 1888) have reported that they woke up frequently during the night and recalled dreaming, apparently as a result of their interest in studying dreams. In all these reports, however, there is no information on how often the subjects normally woke up during the night, although the effect seems to have been of considerable magnitude in at least one case (Nelson, 1888). The author (who was his own subject) complained that his sleep was no longer refreshing as a result of waking too frequently, so that he dampened his interest in dreams in order to get more rest! Renneker (1952a, 1952b) reports on a patient who habitually awoke a few moments before the alarm was to go off, supposedly in order to prevent his dreams from becoming anxiety filled, and presents some psychoanalytic speculations about the nature of this behavior. Omwake and Loranz (1933) and Brush (1930) also report data which leaves little doubt that some subjects can awaken at a preset time.

---

[3] J. Elder, personal communication, 1964.

The most recent study in which the process of dreaming was affected was that of Tart (1963). Using posthypnotic suggestion and working intensively with two subjects, it was found that the subjects could be made to awaken at either the beginnings or ends of their Stage 1 dreams, although the accuracy of this behavior was difficult to judge due to difficulties in defining criteria for "perfect" compliance with the suggestions. As in Frobenius' (1927) study, the subjects had no idea as to how they had awakened, and their dreams did not seem to reflect any processes of discrimination or waking. Tart also found that the Stage 1 dream time of one of the subjects could be increased about 20% by means of posthypnotic suggestions to dream all night.

At present, then, posthypnotic suggestion would seem to be the most powerful technique for affecting the process of dreaming which is selective in its effect (although it is applicable only to a minority of subjects). Total sleep deprivation or dream deprivation is more powerful, but nonselective as it disrupts the phenomena we are trying to manipulate. In general the psychological variables (suggestion, hypnosis, laboratory setting, etc.) seem to have only a small effect on the actual amount of time spent in Stage 1 dreaming.

Discrimination during Sleep

Simon and Emmons (1956) played recorded questions and answers to their subjects once every 5 minutes while they slept. The subjects were asked to call out their own names if they heard the question and answer. The subjects showed essentially no responses to the questions and answers and no evidence of having learned them (judged by questioning them in the morning) whenever they showed Non-stage 1 sleep patterns. The results of this study are difficult to interpret in terms of Stage 1 patterns, however, as they were mixed with arousal patterns.

Granda and Hammack (1961) trained sleep-deprived subjects to press a micro-switch on a certain schedule in order to avoid a noxious electrical shock. By avoiding the shock, the subjects could sleep. They found that their subjects could produce the required avoidance behavior in all stages of sleep, without showing signs of EEG arousal. Williams, Morlock, and Morlock (1963) extended this work and found that the subjects could discriminate between several tones, but that they tended not to respond during Stage 1 dreaming unless the negative reinforcement was quite strong.

Li, Jasper, and Henderson (1952) report that a narcoleptic patient of theirs could sleep through the noise made by the experimenter hitting a brass pail right beside her bed, but would awaken immediately if her name was softly spoken. Oswald (1962) carried out two studies of subjects' reactions to meaningful versus nonmeaningful stimuli during sleep. In one study, subjects showed more muscular responses to the sound of their own names and that of another designated name than to other, nonmeaningful names, although this response was usually accompanied by partial EEG arousal. In a second study, a series of first names was played repeatedly to subjects while they slept, but some of the names had been tape recorded backwards, making them meaningless sounds. The subjects showed more EEG responses without waking (*K* complexes) to the meaningful sounds. A similar finding for tones, one of which had been associated with a noxious stimulus, was reported in cats (Rowland, 1957). Berger (1963), however, did not find any difference in frequency of incorporation into the reported dream content of names of persons with whom the subjects had emotional bonds versus emotionally neutral names.

Thus the literature indicates that the sleeping and dreaming subject is not behaviorally inert, but may make, at least, some fairly simple discriminations and responses, sometimes accompanied by momentary awakening, but sometimes within the sleep or dream state itself.

DISCUSSION

There is little doubt that, in light of the past decade's discoveries, the conclusions of many of the earlier studies will be revised as the results of future research come in. And there are certain methodological problems to be solved even with the use of the EEG-REM technique.[4] In spite of these drawbacks, some conclusions may be drawn from the studies herein reviewed.

Sleep is *not* a state of behavioral inertness (a tradition long given implicit support by the extreme paucity of psychological studies of it). On the contrary, not only is a sleeping person having vivid experiences, dreams, periodically through the night, but (*a*) he is capable of incorporating some external stimuli into the content of his dreams, and (*b*) at least some subjects are capable of sometimes altering their state of consciousness from sleeping (or dreaming) to waking, either by means of a discriminative response to external stimuli or by using the occurrence of dreaming itself or a given time of night as a stimulus. Furthermore, at least some sleeping subjects are able to carry out some relatively simple motor acts in response to external stimuli.

At present then, we know only that some subjects can make responses during sleep, make discriminations, incorporate stimuli, etc. But it should be emphasized that the exact percentages of subjects, the exact behaviors observed, or the methodological difficulties of these studies are not the most important things to be noted in this review. Rather, the fact that *any* subjects can carry out these sorts of behaviors during sleep and dreaming, no matter how imperfectly at present, is the important thing, for it is highly probable that what has been done imperfectly by some subjects, using crude, exploratory methods, can be done far better by other subjects once these techniques are refined and developed. The immediate research need is not to show simply that some subjects can have their sleep and dreaming affected in these manners, but to investigate how well subjects can be trained to do these things, and what the maximal level of performance and effect is. Then we shall possess a number of techniques for actively manipulating dreaming, leading to results which can only be speculated about now (Tart, 1964a), but which will almost certainly surprise us.

The most immediate need in this area is for extensive studies of stimulus incorporation, and of the potentialities for discriminative responses to stimuli, including experiential discriminations within the dream, waking from sleep at preset times, and overt motor responses. How do various sensory modalities differ in the way in which stimuli are incorporated into the dream? Are stimuli in some modalities more distorted than in others? Can the degree of direct incorporation, with distortion, be affected by various experimental techniques? Can subjects learn to incorporate some stimuli effectively enough so that one could use these stimuli as "signals" to indicate to the subject that he should carry out a specific action, for example, dream about a particular topic, or carry out some motor act? What sorts of motor acts can be carried out? There are simple ones, such as raising a finger to indicate that he is dreaming, or that he is experiencing anger within the dream, but what about the possibility of more complex motor acts, say, automatic writing, or sleep talking? Ultimately, to what extent could a "two-way communication system" be developed whereby the experimenter could instruct the subject to do such and such while he is dreaming, and the subject could report on the events of the dream while they are occurring? If such a development were possible, dreams would lose their status as a purely subjective event that can only be reported in retrospect and become a much more immediate sort of behavior.

Besides these questions on what sorts of things can be made to happen, there are also many problems centering around the question of the best method or methods to bring these things about. Judging from the present state of the literature, posthypnotic suggestion seems to offer the most promise, but neither this method nor any of the others have really been adequately tested and explored. What potentialities are there in operant conditioning techniques, effects of increasing centive, various combinations of drugs, etc.? What about combinations of these?

---

[4] In Tart's (1963) study, where subjects were supposed to wake from either the beginnings or ends of their dreams in accordance with posthypnotic suggestions to that effect, difficulties arose in evaluating the results because of a number of responses in which the subjects seemed to be complying on an experiential level with the suggestions, but could not meet strict criteria for the beginnings and ends of their dreams in terms of the amount of time Stage 1 sleep had lasted.

The studies reviewed herein, then, do not sum up our state of knowledge about this area so much as they serve a far more important function, namely, raising questions and pointing out some fascinating possibilities for future research and development.

REFERENCES

Agnew, H., Webb, W., & Williams, R. The effects of stage four sleep deprivation, in press.

Aserinsky, E., & Kleitman, N. Regularly occurring periods of eye motility and concomitant phenomena during sleep. *Science,* 1953, **118**, 273–274.

Aserinsky, E., & Kleitman, N. Two types of ocular motility occurring in sleep. *Journal of Applied Physiology,* 1955, **8**, 1–10.

Barber, T. X. Toward a theory of "hypnotic" behavior: The "hypnotically induced dream." *Journal of Nervous and Mental Disease,* 1962, **135**, 206–221.

Benussi, V. Zur experimentellen Grundlegung hypnosuggestiver Methoden psychischer Analyse. *Psychologische Forschung,* 1927, **9**, 197–274.

Berger, R. Tonus of extrinsic laryngeal muscles during sleep and dreaming. *Science,* 1961, **134**, 840.

Berger, R. Experimental modification of dream content by meaningful verbal stimuli. *British Journal of Psychiatry,* 1963, **109**, 722–740.

Berger, R., & Oswald, I. Eye movement during active and passive dreams. *Science,* 1962, **137**, 601. (a)

Berger, R., & Oswald, I. Effects of sleep deprivation on behavior, subsequent sleep, and dreaming. *Electroencephalography and Clinical Neurophysiology,* 1962, **14**, 297. (b)

Brauchi, J., & West, L. Sleep deprivation. *Journal of the American Medical Association,* 1959, **171**, 11–14.

Brush, E. Observations on the temporal judgement during sleep. *American Journal of Psychology,* 1930, **42**, 408–411.

Calkins, Mary W. Statistics of dreams. *American Journal of Psychology,* 1893, **5**, 311–343.

Cubberly, A. The effects of tensions of the body surface upon the normal dream. *British Journal of Psychology,* 1923, **13**, 243–265.

Dement, W. Dream recall and eye movements during sleep in schizophrenics and normals. *Journal of Nervous and Mental Disease,* 1955, **122**, 263–269.

Dement, W. The effect of dream deprivation. *Science,* 1960, **131**, 1705–1707.

Dement, W., Kahn, E., & Roffwarg, H. The influence of the laboratory situation on the dreams of the experimental subject. *Journal of Nervous and Mental Disease,* in press.

Dement, W., & Kleitman, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology,* 1957, **53**, 339–346. (a)

Dement, W., & Kleitman, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology,* 1957, **9**, 673–690. (b)

Dement, W., & Wolpert, E. The relation of eye movements, body motility, and external stimuli to dream content. *Journal of Experimental Psychology,* 1958, **55**, 543–553.

De Sanctis, S., & Neyroz, U. Experimental investigation concerning the depth of sleep. *Psychological Review,* 1902, **9**, 254–282.

Dittborn, J. Experimental recollection of dreams. *Journal of Psychology,* 1963, **55**, 39–41.

Ehrenwald, H. Versuche zur Zeitauffassung des Unbewussten. *Archiv für die gesamte Psychologie,* 1923, **45**, 144–156.

Elder, J. A study of the ability to awaken at assigned hours. *Psychological Bulletin,* 1941, **38**, 693.

Ellis, H. *The world of dreams.* London: Constable, 1911.

Fisher, C. Studies on the nature of suggestion. Part I: Experimental induction of dreams by direct suggestion. *Journal of the American Psychoanalytic Association,* 1953, **1**, 222–255.

Fisher, C. Dreams and perception: The role of preconscious and primary modes of perception in dream formation. *Journal of the American Psychoanalytic Association,* 1954, **2**, 389–445.

Fisher, C. Subliminal and supraliminal influences on dreams. *American Journal of Psychiatry,* 1960, **116**, 1009–1017.

Foulkes, W. Dream reports from different stages of sleep. *Journal of Abnormal and Social Psychology,* 1962, **65**, 14–25.

Freud, S. *The interpretation of dreams.* (Trans. by J. Strachey) London: Allen & Unwin, 1954.

Frobenius, K. Uber die zeitliche Orientierung im Schlaf und einige Aufwachphänomene. *Zeitschrift für die gasamte Psychologie,* 1927, **103**, 100–110.

Granda, A., & Hammack, J. Operant behavior during sleep. *Science,* 1961, **133**, 1485–1486.

Gresham, S., Webb, W., & Williams, R. Alcohol and caffeine: Effect on inferred visual dreaming. *Science,* 1963, **140**, 1226–1227.

Hawkins, D., Puryear, H., Wallace, C., Deal, W., & Thomas, E. Basal skin resistance during sleep and "dreaming." *Science,* 1962, **136**, 321–322.

Kamiya, J. Behavioral, subjective, and physiological aspects of drowsiness and sleep. In D. Fiske & S. Maddi (Eds.), *Functions of varied experience.* Homewood, Ill.: Dorsey Press, 1961. Pp. 145–174.

Klein, D. B. The experimental production of dreams during hypnosis. *University of Texas Bulletin,* 1930, No. 3009.

Kleitman, N. Patterns of dreaming. *Scientific American,* 1960, **203**, 81–88.

Kleitman, N. *Sleep and wakefulness.* Chicago: Univer. Chicago Press, 1963.

Ladd, G. T. Contribution to the psychology of visual dreams. *Mind,* 1892, **1**(New series), 299–304.

Lesse, S. Experimental studies on the relationship between anxiety, dreams, and dream-like states.

American Journal of Psychotherapy, 1959, **13,** 440–455.

LHERMITTE, J. *Le Sommeil.* Paris: Librarie Armand Colin, 1931.

LI, C., JASPER, H., & HENDERSON, L., JR. The effect of arousal mechanisms on various forms of abnormality in the electroencephalogram. *Electroencephalography and Clinical Neurophysiology,* 1952, **4,** 513–526.

LUBY, E., FROHMAN, C., GRISELL, J., LENZO, J., & GOTTLIEB, J. Sleep deprivation: Effects on behavior, thinking, motor performance, and biological energy transfer systems. *Psychosomatic Medicine,* 1960, **22,** 182–192.

LUBY, E., GRISELL, J., FROHMAN, C., LEES, H., COHEN, B., & GOTTLIEB, J. Biochemical, psychological, and behavioral responses to sleep deprivation. *Annals of the New York Academy of Sciences,* 1962, **96,** 71–78.

MALAMUD, W., & LINDER, F. Dreams and their relationship to recent impressions. *Archives of Neurology and Psychiatry,* 1931, **25,** 1081–1099.

MAX, L. W. An experimental study of the motor theory of consciousness. III: Action-current responses in deaf-mutes during sleep, sensory stimulation, and dreams. *Journal of Comparative Psychology,* 1935, **19,** 469–487.

MIRSKY, A., & CARDON, P. A comparison of the behavioral and physiological changes accompanying sleep deprivation and chloropromazine administration in man. *Electroencephalography and Clinical Neurophysiology,* 1962, **14,** 1–10.

MONROE, W. A study of taste dreams. *American Journal of Psychology,* 1899, **10,** 326–327.

NACHMANSOHN, M. Ueber experimentell erzeugte Traeume nebst kritischen Bemerkungen ueber die psychoanalytische Methodik. *Zeitschrift für die gasamte Neurologie und Psychiatrie,* 1925, **98,** 556–586. In D. Rapaport (Ed.), *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951. Pp. 257–287.

NELSON, J. A study of dreams. *American Journal of Psychology,* 1888, **1,** 367–401.

NEWMAN, R., KATZ, J., & RUBENSTEIN, R. The experimental situation as a determinant of hypnotic dreams. *Psychiatry,* 1960, **23,** 63–73.

OMWAKE, KATHERINE, & LORANZ, MARGARET. Study of the ability to wake at a specified time. *Journal of Applied Psychology,* 1933, **17,** 468–474.

ORNE, M. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology,* 1959, **58,** 277–299.

ORNE, M. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17,** 776–783.

OSWALD, I. *Sleeping and waking: Physiology and psychology.* New York: Elsevier, 1962.

OSWALD, I. Physiology of sleep accompanying dreaming. In, *Scientific Basis of Medicine: Annual Review,* 1964, 102–124.

OSWALD, I., BERGER, R., JARAMILLO, R., KEDDIE, K., OLLEY, P., & PLUNKETT, G. Melancholia and barbiturates: A controlled EEG, body and eye movement study of sleep. *British Journal of Psychiatry,* 1963, **109,** 66–78.

RAMSEY, G. Studies of dreaming. *Psychological Bulletin,* 1953, **50,** 432–455.

RECHTSCHAFFEN, A., & MARON, L. The effect of amphetamine on the sleep cycle. *Electroencephalography and Clinical Neurophysiology,* 1964, **16,** 438-445.

RECHTSCHAFFEN, A., & VERDONE, P. Amount of dreaming: Effect of incentive, adaptation to laboratory, and individual differences. *Perceptual and Motor Skills,* 1964, **19,** 947–958.

RECHTSCHAFFEN, A., VERDONE, P., & WHEATON, JOY. Reports of mental activity during sleep. *Canadian Journal of Psychiatry,* 1963, **8,** 409–414.

RENNEKER, R. Dream timing. *Psychoanalytic Quarterly,* 1952, **21,** 81–91. (a)

RENNEKER, R. Presleep mechanisms of dream control. *Psychoanalytic Quarterly,* 1952, **21,** 528–536. (b)

RENSHAW, S., MILLER, V., & MARQUIS, DOROTHY. *Children's sleep.* New York: Macmillan, 1933.

RODIN, E., LUBY, E., & GOTTLIEB, J. The electroencephalogram during prolonged experimental sleep deprivation. *Electroencephalography and Clinical Neurophysiology,* 1962, **14,** 544–551.

ROFFWARG, H., DEMENT, W., MUZIO, J., & FISHER, C. Dream imagery: Relationship to rapid eye movements of sleep. *Archives of General Psychiatry,* 1962, **7,** 235–258.

ROWLAND, V. Differential electroencephalographic responses to conditioned auditory stimuli in arousal from sleep. *Electroencephalography and Clinical Neurophysiology,* 1957, **9,** 585–594.

SCHATZMAN, M. *Rêves et hallucinations.* Paris: Vigo Frères, 1925.

SCHIFF, S., BUNNEY, W., & FREEDMAN, D. A study of ocular movements in hypnotically induced dreams. *Journal of Nervous and Mental Disease,* 1961, **133,** 59–68.

SCHRÖTTER, K. Experimentelle Träume. *Zentralblatt für Psychoanalyse,* 1912, **2,** 638–648. In D. Rapaport (Ed.), *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951. Pp. 234–248.

SIMON, C. W. Some immediate effects of drowsiness and sleep on normal human performance. *Human Factors,* 1961, March, 1–17.

SIMON, C. W., & EMMONS, W. Responses to material presented during various levels of sleep. *Journal of Experimental Psychology,* 1956, **51,** 89–97.

SNYDER, F. The new biology of dreaming. *Archives of General Psychiatry,* 1963, **8,** 381–391.

SPITTA, H. *Die Schlaf- und Traumzustände der menschlichen Seele.* Tübingen: Fuss, 1882.

STOYVA, J. The effect of suggested dreams on the length of rapid eye movement periods. Unpublished doctoral dissertation, University of Chicago, 1961.

SUCKLING, E., KOENING, E., HOFFMAN, B., & BROOKS, C. McC. The physiological effects of sleeping on hard or soft beds. *Human Biology,* 1957, **29,** 274-288.

SULLY, J. *Illusions: A psychological study*. London: Kegan Paul, Trench, & Trübner, 1905.

TART, C. Frequency of dream recall and some personality measures. *Journal of Consulting Psychology*, 1962, **26**, 467–470.

TART, C. Effects of posthypnotic suggestion on the process of dreaming. Unpublished doctoral dissertation, University of North Carolina, 1963.

TART, C. Hypnotic suggestion as a technique for the control of dreaming. Paper read at American Psychological Association, Los Angeles, 1964. (a)

TART, C. A comparison of suggested dreams occurring in hypnosis and sleep. *International Journal of Clinical and Experimental Hypnosis*, 1964, **12**, 263–289. (b)

TART, C. The hypnotic dream: Methodological considerations and a review of the literature. *Psychological Bulletin*, 1965, **63**, 87–99.

THOMPSON, E. An inquiry into some questions connected with imagery in dreams. *British Journal of Psychology*, 1914, **7**, 300–318.

TITCHENER, E. B. Taste dreams. *American Journal of Psychology*, 1895, **6**, 505–509.

VOLD, J. M. *Über den Traum*. (Ed. by O. Klem) Leipzig: Barth, 1912.

WALSH, W. S. *The psychology of dreams*. New York: Dodd, Mead, 1920.

WEST, L., JANSZEN, H., LESTER, B., & CORNELISOON, F. The psychosis of sleep deprivation. *Annals of the New York Academy of Sciences*, 1962, **96**, 66–70.

WHITMAN, R., PIERCE, C., & MAAS, J. Drugs and dreams. In L. Uhr & J. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960. Pp. 591–595.

WHITMAN, R., PIERCE, C., MAAS, J., & BALDRIDGE, B. Drugs and dreams. II: Imipramine and prochlorperazine. *Comprehensive Psychiatry*, 1961, **2**, 219–226.

WHITMAN, R., PIERCE, C., MAAS, J., & BALDRIDGE, B. The dreams of the experimental subject. *Journal of Nervous and Mental Disease*, 1962, **134**, 431–439.

WILLIAMS, H., GRANDA, A., JONES, R., LUBIN, A., & ARMINGTON, J. EEG frequency and finger pulse volume as predictors of reaction time during sleep loss. *Electroencephalography and Clinical Neurophysiology*, 1962, **14**, 64–70.

WILLIAMS, H., HAMMACK, J., DALY, R., DEMENT, W., & LUBIN, A. Responses to auditory stimulation, sleep loss, and the EEG stages of sleep. *Electroencephalography and Clinical Neurophysiology*, 1964, **16**, 269–279.

WILLIAMS, H., LUBIN, A., & GOODNOW, J. Impaired performance with acute sleep loss. *Psychological Monographs*, 1959, **73**(1, Whole No. 484).

WILLIAMS, H., MORLOCK, H., & MORLOCK, J. Discriminative responses to auditory signals during sleep. Paper read at American Psychological Association, Philadelphia, 1963.

WILLIAMS, H., MORRIS, G., & LUBIN, A. Illusions, hallucinations and sleep loss. In L. J. West (Ed.), *Hallucinations*. New York: Grune & Stratton, 1962. Pp. 158–165.

WILLIAMS, H., TEPAS, D., & MORLOCK, H. Evoked responses to clicks and electroencephalographic stages of sleep in man. *Science*, 1962, **138**, 685–686.

WOOD, P. Dreaming and social isolation. Unpublished doctoral dissertation, University of North Carolina, 1962.

(Received January 4, 1965)

Psychological Bulletin
1965, Vol. 64, No. 2, 92–107

A GENERAL MODEL FOR THE STUDY OF DEVELOPMENTAL PROBLEMS [1]

K. WARNER SCHAIE

West Virginia University

Various research designs employed in developmental psychology for the investigation of maturational and aging effects are examined. It is found that discrepancies and contradictions in the conclusions derived from cross-sectional and longitudinal studies are consequences of the violation of assumptions implicit in these research designs. The conventional methods are shown to be special cases of a general model for research on behavior change over time. The properties of the general model are explicated and the assumptions for the customary designs are reviewed in the light of these properties. The complete model requires consideration of the components of age, time and cohort differences in the identification of developmental change. Both the longitudinal and cross-sectional methods in this context require strong assumptions which can rarely be met. New research strategies are therefore suggested which involve the optimal combination of cross-sectional and longitudinal methods into sequential designs. Such designs and their assumptions are presented both for the independent random sampling and the repeated measurement cases.

It is the purpose of this paper to examine various research designs in developmental psychology with a view to resolving some of the contradictory findings reported in studies employing different designs for the investigation of aging effects. In the course of this examination it is proposed to identify certain inherent inadequacies of the conventional longitudinal and cross-sectional approaches. It will be shown that these methods are merely special cases of a general model for research on behavior changes over time. The underlying assumptions of this model will be explicated and several new and more powerful research strategies will be developed.

Kessen (1960), in his excellent discussion of research designs for developmental problems, suggested that the investigation of simple age functions is naive and proposed that most meaningful designs involve either the interaction of age and environmental changes or the analysis of age functions in special populations. Only in passing did he mention the even more typical case where aging interacts simultaneously with differences in population and environment. Special cases may, of course, be derived where one or the other dimension is ignored. While such analyses may be useful and perhaps often present the only available practical approaches, their limitation must be noted and their relation to the general model well understood.[2]

Kessen proposed that in any developmental problem age-function relationships can be established by cross-sectional study in instances where it is possible to segregate subpopulations across the age range being investigated. He also suggested, however, that studies of variables showing small but reliable changes, the analysis of sequences and of rates of change, and the follow-up of early behaviors which are not found in the repertory of older organisms will usually require the longitudinal approach. It will be shown that such distinctions do *not* apply to the general case and that their application to special cases has often been the result of misunderstood assumptions.

What then are the attributes of a general developmental model from which both cross-sectional and longitudinal methods can be derived and from which their relationship can be clarified? A beginning may be made with Kessen's formulation that "a characteristic is said to be developmental if it can be related to age in an orderly or lawful way." Proceeding further, Kessen's most general definition of a response in a developmental model will be accepted. In this context he specified that a response may be seen as being a function of age, special population, and environment. Such a formulation will be made even more general, however, by recasting its content in terms which will be found more fruitful for further analysis.

---

[1] Parts of this paper were read at a symposium on Research Design in Developmental Psychology at the Los Angeles meetings of the American Psychological Association, September 1964. Preparation of this paper was supported in part by a Public Health Service research grant No. HD 00367-02 from the National Institute of Child Health and Human Development.

[2] Horst (1963, pp. 106–107) discussed in passing the need for a "trapezoid" model which would provide a mathematical basis for our presentation. No such model, however, has been worked out at this date.

The component of "age" will, of course, remain the central ingredient, but it will be more precisely defined to mean the age of the organism at the time of occurrence of the response to be measured. Even more precisely, age will refer to the number of time units elapsed between the birth (entrance into the environment) of the organism and the point in time at which the response is recorded.

The component of "special population" will be replaced by the term "cohort." This term will imply the total population of organisms born at the same point or interval in time. Restrictions as to the nature of the population and the latitude in defining the interval in time designated as being common to a given cohort will, of course, be introduced by the special assumptions appropriate to the problem to be investigated.

The term "environment" will be replaced by the concept of *time of measurement*. In a most general sense the investigator is not really concerned with describing the nature of the total environment but rather with indexing the total environmental impact as occurring at a given temporal point. Such an index of the state of the environment is clearly provided by indicating the time of measurement (see Kuhlen, 1940).

Our general developmental model will now hold that a response is a function of the age of the organism, the cohort to which the organism belongs, and the time at which measurement occurs. It will shortly be seen that these three components are by no means independent. Their confounding, in fact, dictates the methods available for their separate and joint analysis.

Consider next the temporal boundaries of the general developmental model. If the human life-span is assumed to be bounded by a 100-year limit, one can specify the total time required to measure all cohorts at all ages. The time required is simply twice the absolute life-span of the populations under study. Table 1 gives an example, where cohorts have 20-year intervals and measurements are taken also at 20-year intervals, for a total developmental study of a population existing in 1960. It will be seen that this model permits, over its total 200-year span, several independent longitudinal studies but only one cross-sectional study which includes members of all cohorts under investigation. Note that only half the model can be studied unless retro-

TABLE 1

Ages of 20-Year Cohorts at 20-Year Intervals for Cohorts Available to an Investigator in 1960

Time of birth	Age										
1860	0	20	40	60	80	100	—	—	—	—	—
1880	—	0	20	40	60	80	100	—	—	—	—
1900	—	—	0	20	40	60	80	100	—	—	—
1920	—	—	—	0	20	40	60	80	100	—	—
1940	—	—	—	—	0	20	40	60	80	100	—
1960	—	—	—	—	—	0	20	40	60	80	100
Time of measurement	1860	1880	1900	1920	1940	1960	1980	2000	2020	2040	2060

spective measures are available, or unless a future study is planned to begin with the cohort born at the first time of measurement.

Although the complete model can be studied with short-lived organisms, it will generally be most interesting to work with all the cohorts available at any given point of measurement. A general developmental model will therefore be described which will hold for that part of the complete life-span of the population which we wish to investigate. To give this model general applicability it is required that each age level be investigated for each cohort and at each time of measurement within the confines of the model.

GENERAL DEVELOPMENTAL MODEL

An example of a complete model is depicted in Table 2, which gives the ages of cohorts studied over the age range from 5 to 9 years, measured at annual intervals. In general, if the intervals at which a given characteristic is to be measured are called times of measurement (T), then 2T − 1 cohorts with interval latitude equal to the time interval will be required for a study of the complete model.

Let us assume for the moment that the characteristics to be investigated may be found only at the times of measurement and for the cohorts (Cs) listed in Table 2. Let us also assume that scores for an independent random sample of each of the cohorts are available at each of the measurement times. Let us further assume that all groups have had equal amounts of practice on the characteristic to be measured. One of the striking features of Table 2 is the fact that it is impossible to measure all cohorts at all ages or to obtain measures for all ages at each time of measurement. The model can, however, be sectioned horizontally (as indicated by the line in Table 2) to obtain a subset containing measures at all ages for all cohorts. Similarly, by vertical sectioning a subset can be obtained for which all ages are measured at all times of measurement. This means that age is always confounded with either differences between cohorts or with differences between times of measurements. In the general model, it is possible to estimate age changes confounded in either manner, depending upon the question we are interested in, or, as will be shown, one can obtain estimates of net age differences whenever certain assumptions are tenable.

Examining Table 2, it can readily be seen that differences in mean scores between columns in the vertically sectioned subset will represent differences in time of measurement, that is, differences due to commonly experienced environmental variables or changes in cultural context. Likewise, the differences in row means in the horizontally sectioned subset will represent differences between cohorts; that is, the effect of different life experiences previous to the time of measurement or ge-

TABLE 2

AGES OF COHORTS MEASURED AT FIVE AGES WITH ANNUAL MEASUREMENT INTERVALS

Time of birth	Ages								
1951	9	—	—	—	—	—	—	—	—
1952	8	9	—	—	—	—	—	—	—
1953	7	8	9	—	—	—	—	—	—
1954	6_e	7	8	9	—	—	—	—	—
1955	5_f	6_g	7	8	9	—	—	—	—
1956	—	5_h	6_i	7_k	8	9	—	—	—
1957	—	—	5_j	6_l	7	8	9	—	—
1958	—	—	—	5_a	6_b	7	8	9	—
1959	—	—	—	—	5_c	6_d	7	8	9
Time of measurement	1960	1961	1962	1963	1964	1965	1966	1967	1968

netic changes from one cohort to the next. For each subset it is possible also to compare means for the diagonals. Differences between means here involve age differences, which are controlled for time of measurement but confounded with cohort differences in the vertical subset, and controlled for cohort differences but confounded with time of measurement differences in the horizontal subset.

A clearer understanding of the implications of the model may be obtained by deriving the various methods therefrom for the study of developmental change. These include the conventional cross-sectional and longitudinal approaches as well as a third, frequently overlooked but equally important, method which will be identified as the time-lag method.

Relation of the General Model to Conventional Designs

Each of the first five columns in Table 2 represents an example of the conventional *cross-sectional* method. The hypothesis to be investigated simply asks whether there are differences in a given characteristic for samples drawn from different cohorts (i.e., born at different times) but measured at the same time. This is an important question, particularly for the student of individual differences. Age differences in behavior at a given point in time may require differential procedures regardless of the antecedent conditions responsible for these differences. It must be recognized, however, that age differences in the cross-sectional design are inextricably confounded with cohort differences and that it is therefore unreasonable to expect that the difference between Cohorts i and j at Time k will have anything to do with the difference between the mean score of Cohort i at Time k and its mean score at Time l. The fact that the temporal interval in years of birth between Cohorts i and j is identical to the temporal interval between Times k and l for Cohort i is quite coincidental and irrelevant. It is conceivable that the difference in means between a group of 5- and 10-year-olds could be representative of the difference between the mean for the 5-year-olds at age 5 and the mean at age 10 if there were no shifts in the environment and content of our culture. But man's environment does change, and projections over time based on cross-sectional studies are therefore fraught with great hazard.[3]

Let us next look at the last five rows in Table 2. These rows represent the most general case of the *longitudinal* method, that is, independent random samples of the same cohort drawn at successive times of measurement representing different ages.[4] The hypothesis to be investigated here asks whether there are differences in a given characteristic for samples from the same cohort measured at different times. This question is concerned with the effect of age differences within the same population and must be asked whenever one is concerned with predicting age differences in behavior over time. Longitudinal age differences, however, are obviously confounded with time differences. In cross-sectional studies, great risks are taken if one attempts to generalize age differences from one time of measurement to another. In the longitudinal design, it is for the same reasons equally risky to generalize age differences over time from one cohort to another. Personality theorists interested in developmental problems have traditionally been willing to take this risk, typically in the special case where the same sample is measured repeatedly. It will be shown, however, that such risk is a needless one which can be avoided by more efficient designs.

A third method of developmental analysis is suggested also by the general model. It is not generally recognized but should be of particular interest to the social psychologist. This is the method of *time-lag* illustrated by the diagonals in Table 2. The hypothesis here is concerned with asking whether there are differences in a given characteristic for samples of equal age but drawn from different cohorts measured at different times. This is a research design which should be used whenever the investigator is interested in the effect of cultural change upon the performance of individuals of similar age. Again, this effect cannot be measured directly due to its confounding with cohort differences.

---

[3] Sampling problems pose additional issues which must be dealt with as soon as one proceeds beyond the general case. An extended discussion of sampling problems and their management for the cross-sectional method has been given elsewhere (Schaie, 1959).

[4] It should be noted that the conventional longitudinal method is simply a special case involving the repeated measurement of the same sample from a given cohort. The advantages and liabilities of the repeated measurement case will be considered in a later section of this paper.

Methods for Separating Sources of Developmental Change

It will now be shown that the general model may be used to generate methods which make it possible, given certain assumptions, to separate the three sources of developmental change. Several definitions will be required for this purpose:

1. Cohort difference (Cd) = the net effect of differences between a younger cohort, C_i, and an older cohort, C_j, for random samples of subjects controlled for the effect of age and time of measurement; that is, net changes between generations.
2. Time difference (Td) = the net effect of differences between an earlier time of measurement, T_k, and a later time of measurement, T_l, for random samples of subjects, controlled for the effect of age and cohorts; that is, net changes within the environment.
3. Age difference (Ad) = the net effect of maturational change (over time) from a younger age, A_m, to an older age, A_n, between random samples of subjects controlled for the effect of cohorts and times of measurement; that is, net age changes within the organism.

Next, let us define the three methods of measuring change over time, which we have deduced from the general model, as consisting of the following additive components: (*a*) Cross-sectional difference (CSd) = age differences plus cohort differences (CSd = Ad + Cd); (*b*) Longitudinal difference (LOd) = age differences plus time differences (LOd = Ad + Td); (*c*) Time-lag difference (TLd) = time differences plus cohort differences (TLd = Td + Cd). Transposing and changing signs, it can be shown that:

$$\mathrm{Ad} = \mathrm{CSd} - \mathrm{Cd}, \qquad [1]$$

$$\mathrm{Cd} = \mathrm{TLd} - \mathrm{Td}, \qquad [2]$$

$$\mathrm{Td} = \mathrm{LOd} - \mathrm{Ad}. \qquad [3]$$

It can also be shown by substitution that the following relationships will hold: Substituting Formula 3 into Formula 2, and Formula 2 into Formula 1 will yield

$$\mathrm{Ad} = \mathrm{CSd} - [\mathrm{TLd} - (\mathrm{LOd} - \mathrm{Ad})]. \qquad [4]$$

Removing parentheses, collecting terms, and solving for *age difference*, it is then found that

$$\mathrm{Ad} = \frac{\mathrm{CSd} - \mathrm{TLd} + \mathrm{LOd}}{2}. \qquad [5]$$

Similarly, substituting Formula 1 into Formula 3, and Formula 3 into Formula 2, we obtain

$$\mathrm{Cd} = \mathrm{TLd} - [\mathrm{LOd} - (\mathrm{CSd} - \mathrm{Cd})]. \qquad [6]$$

Again removing parentheses, collecting terms, and solving for *cohort difference*, it is found that

$$Cd = \frac{TLd - LOd + CSd}{2}. \qquad [7]$$

Finally, substituting Formula 2 into Formula 1, and Formula 1 into Formula 3, it may be found that

$$\mathrm{Td} = \mathrm{LOd} - [\mathrm{CSd} - (\mathrm{TLd} - \mathrm{Td})]. \qquad [8]$$

Solving Formula 8 for *time difference*, it is then shown that

$$\mathrm{Td} = \frac{\mathrm{LOd} - \mathrm{CSd} + \mathrm{TLd}}{2}. \qquad [9]$$

Unfortunately, it is not possible to compute these net effects for the general model as a whole, since equal representation cannot be obtained for all parameters. As indicated above, one can either measure all cohorts at all possible measurement times with unequal representation of ages, or one can measure all cohorts at all ages with unequal representation of measurement dates; or lastly, all ages can be measured at all measurement times with unequal representation of cohorts. This restriction is not severe, however, since it is in most cases virtually impossible to obtain successive independent random samples over the life span of all cohorts in a given population.

While it is impossible to separate the sources of variance associated with developmental change for the total model, such unconfounding can readily be effected for any set of samples for which independent estimates of each of the three developmental components are available. With respect to age and cohort differences this will be true for any subset or subsets containing a minimum of two cohorts; independent random samples from which are measured at two ages. It can also be accomplished with respect to age and time differences for subsets with a minimum of two ages; independent random samples of cohorts representing these ages being measured at two times. Moreover, unconfounding can be effected with respect to time and cohort differences for subsets consisting of at least two cohorts which are measured at least twice.

The minimal subset which can be subjected to complete analysis is indicated in the group which is enclosed by a triangle in Table 2, since it includes all the minimal combinations specified above. Depending on the generalization intended, it will therefore usually be most economical either to measure at least two cohorts many times or to measure many cohorts at least twice. If a third cohort or measurement occasion is added, moreover, further checks on one's assumptions become possible.

To summarize, the above considerations show clearly that it becomes possible to unconfound the sources of developmental change as soon as one begins to analyze simultaneously two or more behavior sequences. In view of the obvious advantage of obtaining unconfounded measures of age changes, it follows that sequential strategies should be preferred to the conventional longitudinal and cross-sectional methods in most instances. Appropriate sequential strategies will next be discussed in detail; first, for the case of independent samples, and then for the special case of repeated measurements.

COHORT-SEQUENTIAL METHOD

Whenever the subset of samples to be examined contains measures for all cohorts measured at all ages, it may be concluded that it is the investigator's intention to generalize over cohort differences. Such a design samples cohort differences at many times within the life span of the cohorts. This design will be referred to as the *cohort-sequential* method since longitudinal sequences for two or more cohorts are examined simultaneously. The cohort-sequential method permits inferences as to age changes at all points of the age range covered and, also, inferences about cohort differences at all ages. The two bottom rows of Table 2 are an example of a typical cohort-sequential design.

The entries in Table 2 which have been labeled a, b, c, and d represent the minimum case for this type of analysis. The questions which can be answered in the illustrative minimum case are as follows:

What is the net age change over one time unit (year) at age 5 (as sampled by Cohorts 1958 and 1959)?

Substituting the appropriate cell labels from Table 2 into Formula 5, it may be shown that

$$\mathrm{Ad} = \frac{(c-b)-(c-a)+(c-d)}{2} = \frac{a-b+c-d}{2}.$$

Or, in general, what is the net age change from age m to age n as sampled for cohorts i and j?

$$\mathrm{Adcs} = \frac{\mathrm{C}_i\mathrm{A}_m - \mathrm{C}_i\mathrm{A}_n + \mathrm{C}_j\mathrm{A}_m - \mathrm{C}_j\mathrm{A}_n}{2} \quad [10]$$

The estimate of the net age difference is obtained by averaging age differences for two successive cohorts measured over successive time intervals.

What is the net cohort difference on a given variable between Cohorts 1958 and 1959 (as sampled at ages 5 and 6)?

Substituting into Formula 7, it will be shown that

$$\mathrm{Cd} = \frac{(c-a)-(c-d)+(c-b)}{2} = \frac{c+d-a-b}{2}.$$

Or, in general, what is the net cohort difference between Cohorts i and j, sampled at ages m and n?

$$\mathrm{Cdcs} = \frac{\mathrm{C}_j\mathrm{A}_m + \mathrm{C}_j\mathrm{A}_n - \mathrm{C}_i\mathrm{A}_m - \mathrm{C}_i\mathrm{A}_n}{2} \quad [11]$$

This estimate averages cohort differences at two ages measured over two successive time intervals.

What is the interaction between age and cohort differences (attributable to and confounded with time of measurement differences)?

Substituting into Formula 9, it is found that

$$\text{Td}=\frac{(c-d)-(c-b)+(c-a)}{2}=\frac{b+c-a-d}{2}.$$

Or, in general, what is the average net change per time unit for Cohort i at age m and Cohort j at age n?

$$\text{Td}_{\text{CS}}=\frac{\text{C}_i\text{A}_n+\text{C}_j\text{A}_m-\text{C}_i\text{A}_n-\text{C}_j\text{A}_m}{2} \quad [12]$$

This estimate averages cohort differences irrespective of, and controlling for age level in order to obtain a time difference. It is necessarily confounded as it is impossible to measure two different cohorts of the same age at the same time of measurement.

The analysis of variance may be used to provide a test of significance of the net differences in the cohort-sequential design. Table 3 illustrates the relationships described above in the analysis of variance context. It may be noted that increasing the number of successive times of measurement for all cohorts will extend the age range for which cohort differences can be evaluated, while increasing the number of sequences would augment the number of cohorts over which age differences can be generalized.

The cohort-sequential design is clearly the method of choice when a test of the generalizability of constructs over different generations (cohorts) is required. It will be noted that this design controls for cohort differences which might be attributable to genetic changes at all ages and that the cohort differences are controlled for age differences. But it is not possible in this design to identify an unconfounded time difference component. This means that *the cohort-sequential method will yield unambiguous results only when the assumption is met that changes in the variable under study are unrelated to cultural change.*[5] This assumption is rarely met except in the study of physical attributes or in investigations involving infrahuman organisms. Even when this assumption is met, it is as difficult to obtain an ample coverage of the life span of the organism in this variant as in the single-cohort or repeated measurement longitudinal design. Moreover, it will be seen that the cohort-sequential method will provide no more accurate estimates of age differences than do the more efficient alternative procedures.

TABLE 3

ANALYSIS OF VARIANCE FOR THE COHORT–SEQUENTIAL MODEL

Source of variation	Degrees of freedom
Between cohorts (C)	$C-1$
Between ages (A)	$A-1$
Cohort × age interaction	$(C-1)(A-1)$
Error	$N^{a}-(C)(A)$
Total variation	$N-1$

[a] Number of observations.

TIME-SEQUENTIAL METHOD

The first alternative procedure to be advocated is called the *time-sequential* method and applies whenever the subset of samples under examination contains measures of all ages at all times of measurement. The time-sequential method permits inferences as to age differences at all points of the age range covered, as well as inferences about time differences (or cultural shifts) at all ages. The first two columns of Table 2 give an example of a typical time-sequential design.

The entries in Table 2 which are labeled e, f, g, and h represent the minimum case for time-sequential analysis. The propositions which can be evaluated for the illustrative case are as follows:

What is the net age change over one time unit (year) at age 5 (as sampled at Measurement Times 1960 and 1961)?

Substituting the cell identifiers from Table 2 into Formula 5, it is found that

$$\text{Ad}=\frac{(f-e)-(f-h)+(f-g)}{2}=\frac{f+h-e-g}{2}.$$

Or, in general, the net age change from age m to age n as sampled at Times k and l is given by

$$\text{Ad}_{\text{TS}}=\frac{\text{T}_k\text{A}_m+\text{T}_l\text{A}_m-\text{T}_k\text{A}_n-\text{T}_l\text{A}_n}{2}. \quad [13]$$

---

[5] The conventional longitudinal method as a special case of the cohort-sequential method is encumbered by the same restrictive assumption.

The estimate of age differences is obtained by averaging age differences at two successive times of measurement for two successive cohort pairs.

What is the net time-lag (cultural shift) for one time unit (year) at Time 1960 (as sampled at ages 5 and 6)?

Substituting into Formula 9, it will be shown that

$$\mathrm{Td} = \frac{(f-g)-(f-e)+(f-h)}{2} = \frac{e+f-g-h}{2}.$$

Or, in general, the net time-lag difference from Time k to Time l as sampled at ages m and n is

$$\mathrm{Td_{TS}} = \frac{\mathrm{T}_k\mathrm{A}_n + \mathrm{T}_k\mathrm{A}_m - \mathrm{T}_l\mathrm{A}_n - \mathrm{T}_l\mathrm{A}_m}{2}. \quad [14]$$

This estimate averages the time lag differences at two ages for two successive pairs of cohorts.

What is the interaction between age and time-lag differences (attributable to and confounded with cohort differences)?

Substituting into Formula 7, it is found that

$$\mathrm{Cd} = \frac{(f-h)-(f-g)+(f-e)}{2} = \frac{f+g-e-h}{2}.$$

Or, in general, the average net difference between cohorts measured at Time k for age m and Time l for age n is

$$\mathrm{Cd_{CS}} = \frac{\mathrm{T}_k\mathrm{A}_m + \mathrm{T}_l\mathrm{A}_n - \mathrm{T}_k\mathrm{A}_n - \mathrm{T}_l\mathrm{A}_m}{2}. \quad [15]$$

This estimate averages age differences for successive times of measurement to obtain a cohort difference. It is confounded, as it is impossible to obtain members of the same cohort who retain the same age at two different times of measurement.

The analysis of variance is again the appropriate technique for the evaluation of differences in the time-sequential design. Table 4 gives the analysis of variance design for the time-sequential model. In this design, adding to the number of ages measured at each time of measurement will extend the age range for which time-lag differences can be evaluated. Increasing the number of times of measurement would serve to augment the degrees of freedom available for evaluating overall time-lag differences.

TABLE 4

Analysis of Variance for the Time-Sequential Model

Source of variation	Degrees of freedom
Between times (T)	$T-1$
Between ages (A)	$A-1$
Time × age interaction	$(T-1)(A-1)$
Error	$N-(T)(A)$
Total variation	$N-1$

Two measurements of every cohort will suffice to permit the evaluation of time-lag at all ages of interest to the investigator. It would seem, therefore, that this method should be favored whenever generalizability of a construct over time is at issue. Note that the time-sequential model controls age differences for cultural shift and time-lag differences for the effect of age. It is impossible, however, to identify an unconfounded cohort difference component. This means that the *time-sequential method will yield unambiguous results only when the assumption is met that change in the variable under study is unrelated to genetic or cohort-specific environmental changes.*[6] This assumption is a hazardous one only when physical attributes are involved or when behaviors are studied which are thought to be affected by gross environmental crises such as wars or severe depressions. The assumption would seem to be quite tenable for many variables in the areas of ability, personality, learning, interests, and, in fact, in most matters of concern to the developmental psychologist.

If neither of the assumptions underlying the time-sequential or cohort-sequential model can be met, then recourse must be taken to a third sequential model.

CROSS-SEQUENTIAL METHOD

The cross-sequential method may be applied whenever a subset of samples contains measures of all cohorts at all times of measurement. The cross-sequential method permits inferences as to time-lag differences at all points of the time range covered as well as to cohort differences at all times of measurement. Any rectangular area of Table 2 may be subjected to cross-sequential analysis.

---

[6] The assumptions given for the time-sequential model also apply to the conventional cross-sectional method.

The entries of Table 2 which are labeled i, j, k, and l represent the minimum subset for cross-sequential analysis. The propositions which can be evaluated for the illustrative case are as follows:

What is the net cohort difference on a given variable between Cohorts 1956 and 1957 (as sampled at Measurement Times 1962 and 1963)?

Substituting in Formula 7, it is found that

$$\mathrm{Cd}=\frac{(i-l)-(i-k)+(i-j)}{2}=\frac{i+k-j-l}{2}.$$

Or, in general, the net cohort difference between Cohorts i and j sampled at Times k and l is

$$\mathrm{Cd}_{\mathrm{XS}}=\frac{\mathrm{C}_i\mathrm{T}_k+\mathrm{C}_i\mathrm{T}_l-\mathrm{C}_j\mathrm{T}_k-\mathrm{C}_j\mathrm{T}_l}{2}. \quad [16]$$

This estimate averages cohort differences at two times of measurement over two successive age slices.

What is the net time-lag (cultural shift) for one time unit (year) at Time 1962 (as sampled for Cohorts 1956 and 1957)?

Substituting into Formula 9, it is found that

$$\mathrm{Td}=\frac{(i-k)-(i-j)+(i-l)}{2}=\frac{i+j-k-l}{2}.$$

Or, in general, the net time difference between Times k and l sampled for Cohorts i and j is

$$\mathrm{Td}_{\mathrm{XS}}=\frac{\mathrm{C}_i\mathrm{T}_k+\mathrm{C}_j\mathrm{T}_k-\mathrm{C}_i\mathrm{T}_l-\mathrm{C}_j\mathrm{T}_l}{2}. \quad [17]$$

This estimate averages time differences for two cohorts over two successive age slices.

What is the interaction between cohort and time-lag differences (attributable to and confounded with age differences)?

Substituting into Formula 5, it is found that

$$\mathrm{Ad}=\frac{(i-j)-(i-l)+(i-k)}{2}=\frac{i+l-j-k}{2}.$$

Or, in general, the average net age increment per age unit for Cohort i at Time k and for Cohort j at time l is

$$\mathrm{Ad}_{\mathrm{XS}}=\frac{\mathrm{C}_i\mathrm{T}_k+\mathrm{C}_j\mathrm{T}_l-\mathrm{C}_j\mathrm{T}_k-\mathrm{C}_i\mathrm{T}_l}{2}. \quad [18]$$

The analysis of variance design for the cross-sequential model is given in Table 5. Adding to the number of cohorts in this design will increase the range over which time differences can be generalized, while augmenting the number of times of measurement will permit generalization over time for the cohort differences. Adding both rows and columns will in addition increase the generalizability of the cohort-time interaction.

The cross-sequential model controls time differences for changes over generations and cohort differences for cultural shift. No unconfounded age difference component is available in this model, and consequently it follows that *time-lag methods will yield unambiguous results only when the assumption is met that change in the variable under study is unrelated to age changes*. This assumption may be tenable for many characteristics which remain stable over substantial portions of the organism's life span. It is not usually supportable in developmental studies of childhood but may frequently apply to investigations involving behavior change over time in adults. It should be noted that the cross-sequential model can be applied to any research design employing either the cohort or time-sequential methods by simply adding one more sample as indicated in the upper triangle in Table 2.

TABLE 5

ANALYSIS OF VARIANCE FOR THE CROSS-SEQUENTIAL MODEL

Source of variation	Degrees of freedom
Between cohorts (C)	$C-1$
Between times (T)	$T-1$
Cohort × time interaction	$(C-1)(T-1)$
Error	$N-(C)(T)$
Total variation	$N-1$

Empirical Checks for the Assumptions for the Sequential Models

Numerous research situations will arise where there is no independent evidence as to the tenability of the assumptions required for unambiguous inferences from the sequential strategies. From the known relations between the components of developmental change, it will now be possible to point to a number of empirical checks of the assumptions. It will also be possible to identify the consequences of known violations of assumptions, since their result will not be random confounding but will lead to combinations of components which may be of inherent interest. The check of assumptions will generally require conducting analyses for two of the models. This requirement is unreasonable for a combination of the cohort-sequential and time-sequential procedures but, as was pointed out, involves only little extra effort for a combination of either with the cross-sequential strategy. Let us now examine each of the eight possible empirical situations.

Response is unrelated to developmental change. This situation may be detected by any of the three strategies and will obtain when both main effects and the interaction are nonsignificant. Since some developmental change is known to operate in almost any variable, there is danger in accepting the null hypothesis. Even where no differences are found in an analysis via two models, one should be cautious in generalizing from such findings unless large samples have been used. Such a lack of relationship will most likely be found only for limited age, time, or cohort spans.

Response is a simple function of age: $R = f(A)$. If this situation is true, Ad estimates from the cohort- and time-sequential models will be equivalent. The Cd term in the cohort-sequential and the Td term in the time-sequential models will be zero. Both Cd and Td terms in the cross-sequential model will be alternate estimates of age change and will equal each other. No interaction can be significant in this instance. Such a finding would mean that there are no genetic or cultural differences, a circumstance most likely to be found in well-controlled animal studies involving acquired behaviors. In human subjects, this outcome implies that the behavior studied varies as a function of a stable maturational sequence.

Response is a simple function of cohort differences: $R = f(C)$. Here the Cd estimates from the cohort- and cross-sequential models will be equivalent. The Ad term in the cohort-sequential and the Td term in the cross-sequential procedures will be zero. Both Ad and Td terms in the time-sequential model will be alternate estimates of cohort differences and will equal each other. Again, no interaction can be significant. These results would mean that there are no cultural or age changes but that response differences must be attributed solely to variation among generations. This situation will arise rarely, except where true genetic changes in physical attributes may be observed.

Response is a simple function of time differences: $R = f(T)$. In this case the Td estimates from the time- and cross-sequential procedures will be equivalent. The Ad term in the time-sequential and the Cd term in the cross-sequential model will be zero. Both Ad and Cd terms in the cohort-sequential model become alternate estimates of time differences which will equal each other. No interaction can be significant in this case. This model implies that there are no age or genetic differences and that response differences must be attributed to time or cultural changes. A situation of this type will probably prevail for behaviors which ordinarily remain stable once they are acquired but which may be modified due to environmental changes. Word usages, attitude patterns, occupational choices, modes of dress, etc., may fit this model.

Response is a function of age and cohort differences: $R = f(A,C)$. If both age and cohort differences are present, the Ad term from the time-sequential model will be significantly larger than the Ad term from the cohort-sequential model. This must be so because the latter is the net Ad estimate, while the former is inflated by cohort differences. Similarly, the Cd term in the cross-sequential model will be significantly larger than the Cd term in the cohort-sequential model. The cohort-sequential Cd term is, of course, the proper Cd estimate in this case. The Td term in the time-sequential model will here actually provide an alternate estimate of net cohort differences, and the Td term in the cross-sequential model is the alternate estimate of net age differences. The interaction term in the cohort-sequential model, if significant, implies differential age changes for different cohorts. The interactions for the other models are confounded and not interpretable. Most motor skills and perceptual variables which have maturational sequences and show some genetic shift, but which are unrelated to cultural change, will fit this paradigm.

Response is a function of age and time differences: $R = f(A,T)$. If developmental change is due to both age and time differences, then the Ad term from the cohort-sequential model must be significantly larger than the Ad term from the time-sequential model. The latter is the true Ad estimate, while the former is inflated by confounding with time differences. Furthermore, the Td term from the cross-sequential model will be significantly larger than the Td term from the time-sequential model. The former is inflated by confounding with age differences and the latter is the net Td estimate. The Cd estimate in the cohort-sequential model in this case provides an alternate estimate of time differences and the Cd estimate in the cross-sequential model estimates the net age differences. The interaction term in the time-sequential model, if significant, may be interpreted to mean differential age changes at different time intervals. The interactions for the other models are again confounded and consequently not interpretable. Most acquired behaviors with maturational sequences will fit this paradigm provided they are unrelated to genetic changes.

Response is a function of cohort and time differences: $R = f(C,T)$. This paradigm is perhaps the most important one for the student of adult age changes, as maturational sequences have often been imputed by researchers where careful analysis would have shown cohort and time differences but no maturational change whatsoever. In this situation the Cd term from the cross-sequential model will be the true Cd estimate, and it will be significantly smaller than the Cd term from the cohort-sequential model. The latter represents the cohort differences plus time differences. Also, the Td term from the cross-sequential model will be significantly smaller than the Td term from the time-sequential model. The former is the net Td estimate, while the latter again represents the cohort difference inflated by time differences. The Ad term from the cohort-sequential model yields an alternate estimate of time differences and the Ad term from the time-sequential model represents net cohort differences. The interaction term in the cross-sequential model, if significant, may be interpreted as differential time differences for different cohorts. The interactions for the other models are again not interpretable. This paradigm will fit many adult behaviors which persist through the individual's life span once they are established at maturity, but which show differences due to genetic and cultural factors. Some of the mental abilities, many personality variables, and certain physical characteristics which are modified by nutritional and other environmental intervenors will probably be found to show this pattern upon proper analysis.

Response is a function of age, cohort and time differences: $R = f(A,C,T)$. The last alternative will probably be found in the total life-span study, or at least in studies involving childhood and adolescence of the kind of variables which in adults will fit the $R = f(C,T)$ paradigm. It will, of course, also apply to the limited number of variables with maturational sequences which show both cultural and genetic changes. In this situation, the Ad term from the cohort-sequential model will be inflated by time differences and the Ad term from the time-sequential model will be inflated by cohort differences. Similarly, in the cross-sequential model both the Td and Cd terms will be inflated by age differences. The Cd term in the cohort-sequential model and the Td term in the time-sequential model both represent cohort differences plus time differences. The interaction in the cohort-sequential model implies differential age changes for different cohorts, and the interaction in the time-sequential model represents different age changes at different times. The cross-sequential interaction, if significant, again represents differential cohort changes at different time periods. It should be noted that the confounded Ad + Cd and Ad + Td terms accurately represent what are conventionally described as cross-sectional and longitudinal age changes.

The nature of the general developmental model as explained previously does not permit complete unconfounding under these circumstances unless additional assumptions about the form of the age gradients are introduced. The investigator who is interested in the determination of net time or cohort differences, however, can always obtain such information by analyzing the age range or age ranges at which little or no maturational change occurs. Practically all organisms show such periods for most variables of interest. Similarly, if two or more successive cohorts can be found which show no genetic change, it then becomes possible to get net Ad and Td estimates. If two or more measurement times are found for which there is little or no cultural shift, then it is also possible to obtain net Ad and Cd estimates.

Table 6 presents a summary of the meaning of the various difference components for each of the above-described paradigms. Inspection of this table shows that the investigator without prior information on the paradigm best fitting his data will be most likely to obtain information on net age differences by use of the cross-sequential method. The best bet for obtaining net cohort differences will be found in the time-sequential method, and coverage for most contingencies in finding net time-lag differences will be obtained by using the cohort-sequential method.

Decision Rules

From the above paradigms, it now becomes possible to generate logical decision rules for the evaluation of empirical data which will permit inferences as to the proper meaning of any developmental changes found in such data. The reader must, however, be cautioned that use of the proposed decision rules is conditional upon the tenability of our assumptions about the slope of the developmental gradients. The scheme presented in Table 6 assumes that all differences are unidirectional and positive; that is, an incremental progression exists from a younger to an older age, from an earlier to a later measurement time, and from a younger to an older generation. The scheme will still work if all gradients are negative, but difficulties arise when one of the developmental components differs in slope. Appropriate modification of the schema and the resulting decision rules will therefore be required whenever any of the components of developmental change departs from the assumptions of unidirectionality and/or linearity.[7]

TABLE 6

INTERPRETATION OF DIFFERENCE COMPONENTS OBTAINED BY THE ANALYSIS OF VARIANCE FOR THE SEQUENTIAL MODELS UNDER THE DIFFERENT ASSUMPTIONS

Paradigm	Cohort-Sequential Model			Cross-Sequential Model			Time-Sequential Model		
	Age diff.	Cohort diff.	Ad×Cd interact.	Cohort diff.	Time-lag diff.	Cd×Td interact.	Time-lag diff.	Age diff.	Td×Ad interact.
R=f(A)	Ad	—	—	Ad	Ad	—	—	Ad	—
R=f(C)	—	Cd	—	Cd	—	—	Cd	Cd	—
R=f(T)	Td	Td	—	—	Td	—	Td	—	—
R=f(A,C)	Ad	Cd	Ad×Cd	Ad+Cd	Ad	—	Cd	Ad+Cd	—
R=f(A,T)	Ad+Td	Td	—	Ad	Ad+Td	—	Td	Ad	Td×Ad
R=f(C,T)	Td	Cd+Td	—	Cd	Td	Cd×Td	Cd+Td	Cd	—
R=f(A,C,T)	Ad+Td	Cd+Td	Ad×Cd	Ad+Cd	Ad+Td	Cd×Td	Cd+Td	Ad+Cd	Td×Ad

Note.—The age-difference column for the cohort-sequential model holds true also for the conventional longitudinal method, while the age-difference column for the time-sequential model applies to the conventional cross-sectional method. All differences are assumed to be unidirectional.

---

[7] Nine basic types of slopes may be recognized. These are: zero, positive, negative, concave, convex, positive asymptotic, negative asymptotic, accelerating, and decelerating. As a result there are as many as 729 different models when the slopes of the three components are jointly considered. In certain non-

Decision rules will now be given for data which are analyzed by both the cross-sequential and time-sequential methods and also for data which are analyzed by both cross-sequential and cohort-sequential methods. It is assumed that few investigators will have occasion to analyze a set of data by both cohort-sequential and time-sequential methods, but similar decision rules for this case can readily be derived from Table 6 if needed.

Decision rules for the TS-XS analyses are as follows:

1. If $Td_{ts} \simeq 0$; then $R = f(A)$; if not, and
2. if $Td_{xs} \simeq 0$; then $R = f(C)$; if not, and
3. if $Cd_{xs} \simeq 0$; then $R = f(T)$; if not, and
4. if $Td_{ts} + Td_{xs} - Ad_{ts} \simeq 0$; then $R = f(A, C)$; if not, and
5. if $Td_{ts} + Cd_{xs} - Td_{xs} \simeq 0$; then $R = f(A, T)$; if not, and
6. if $Td_{xs} + Ad_{ts} - Td_{ts} \simeq 0$; then $R = f(C, T)$; if not, then $R = f(A, C, T)$.

Decision rules for the CS-XS analyses are as follows:

1. If $Cd_{cs} \simeq 0$; then $R = f(A)$; if not, and
2. if $Ad_{cs} \simeq 0$; then $R = f(C)$; if not, and
3. if $Cd_{xs} \simeq 0$; then $R = f(T)$; if not, and
4. if $Cd_{cs} + Td_{xs} - Cd_{xs} \simeq 0$; then $R = f(A, C)$; if not, and
5. if $Cd_{cs} + Cd_{xs} - Ad_{cs} \simeq 0$; then $R = f(A, T)$; if not, and
6. if $Ad_{cs} + Cd_{xs} - Cd_{cs} \simeq 0$; then $R = f(C, T)$; if not, then $R = f(A, C, T)$.

Empirical estimates will, of course, be subject to sampling fluctuations and the definition for zero will have to be subject to the level of confidence specified for a given experiment.[8]

REPEATED MEASUREMENT DESIGNS

Whenever experimental error is large or where differences between experimental conditions, though significant and stable, are expected to be small, variability may be reduced markedly by use of the matched group design. The conventional longitudinal method is essentially an application of such a design. It will now be shown how the matched group design can be applied to the sequential strategies and models discussed above. Before going into such detail, it will be useful to consider some of the general problems of matched group designs in developmental studies.

In a repeated measurement design the investigator obtains measures on the same sample of subjects at different times. Such repetition poses no problems where physical attributes are involved, but complications arise whenever variables are studied which are amenable to practice or where the testing procedure affects the behavior to be measured. Instead of comparing independently derived groups of subjects and being limited thereby to statements of the difference between group means, the repeated measurement design permits comparing each observation of a subject's behavior with some earlier or later measurements. In theory, therefore, it should be possible to detect small but reliable changes because the matching procedure will result in more sensitive estimates of such changes. Unless adequate controls are provided, and these are for practical and understandable reasons usually lacking in developmental studies, the small but significant changes may simply be a function of practice or other artifacts.

It is not difficult to draw new random samples from a carefully defined population over long periods of time. In contrast, it is virtually impossible to maintain a single sample in its entirety over any length of time. Unfortunately, sampling attrition is not random but is often related to the very variables being studied. Unless it can be shown that the reduced sample for which repeated measurements are available is substantially similar to the original sample, it is, of course, not possible to generalize even for the single cohort sampled. No such evidence is readily available for most of the various longitudinal studies reported in the literature.

Attention should further be called to the fact that for an equal number of observations the degrees of freedom available for the error term used to test mean differences will always

---

linear designs, moreover, confounding may not be complete. It is unlikely that there are empirical data which fit each of the possible models. Also, empirical data might equally well fit several of these models. A more detailed analysis of this problem goes beyond the scope of this paper and will be attempted elsewhere.

[8] The decisions will be based on a comparison of the mean squares obtained from the analyses of variance as detailed in Tables 3–5 and 7–9.

be less for the replicated than for independently derived measurements. The gain in sensitivity must, therefore, exceed the loss caused by testing a smaller number of individuals before the repeated measurement design will pay off. The repeated measurement design will always be inferior to the independent sample design in evaluating cohort differences, since such differences must of necessity involve the comparison of independent samples.

Tables 7 and 8 give the analysis of variance designs for the cohort- and cross-sequential models with repeated measurements. A time-sequential model is not available for the repeated measurement case, since repeated measurements would not be available for the first and last cohort represented in each subset (see Table 2 and the discussion of the general model). It will be noted that the significance of cohort differences in both cohort- and cross-sequential models will be tested by an error term based on the residual variation between subjects which in the cohort-sequential model will have $N - (N/\mathrm{A}) - \mathrm{C}\ (\mathrm{A} - 1)$ fewer degrees of freedom than in an independent sample design based on the same number of observations. Similarly, the error term to be used for testing the significance of cohort differences in the cross-sequential model will have $N - (N/\mathrm{T}) - \mathrm{C}\ (\mathrm{T} - 1)$ fewer degrees of freedom in the repeated measurement than in the independent sampling design.

TABLE 7

Analysis of Variance for the Cohort-Sequential Model with Repeated Measurements

Source of variation	Degrees of freedom
Between subjects	$\frac{N}{\mathrm{A}}-1$
Between cohorts	$\mathrm{C}-1$
Residual between subjects	$\frac{N}{\mathrm{A}}-\mathrm{C}$
Within subjects	$N-\frac{N}{\mathrm{A}}$
Between ages	$\mathrm{A}-1$
Cohort × age interaction	$(\mathrm{C}-1)(\mathrm{A}-1)$
Residual within subjects	$N-\left(\frac{N}{\mathrm{A}}\right)-\mathrm{C}(\mathrm{A}-1)$
Total variation	$N-1$

TABLE 8

Analysis of Variance for the Cross-Sequential Model with Repeated Measurements

Source of variation	Degrees of freedom
Between subjects	$\frac{N}{\mathrm{T}}-1$
Between cohorts	$\mathrm{C}-1$
Residual between subjects	$\frac{N}{\mathrm{T}}-\mathrm{C}$
Within subjects	$N-\frac{N}{\mathrm{T}}$
Between times	$\mathrm{T}-1$
Cohort × time interaction	$(\mathrm{C}-1)(\mathrm{T}-1)$
Residual within subjects	$N-\left(\frac{N}{\mathrm{T}}\right)-\mathrm{C}(\mathrm{T}-1)$
Total variation	$N-1$

The age difference and age-by-cohort-interaction terms in the cohort-sequential repeated measurement design will be confounded both with time differences and practice effects. The error term available to test these differences is based on the residual variance within subjects. This error term will have $N - (N/\mathrm{A})$ fewer degrees of freedom than the comparable error term in an independent sample design with the same number of observations. Similarly, in the cross-sequential repeated measurement design, the time-difference and cohort-time interactions will be confounded with age differences and practice effects. The error term available to test these differences will have $N - (N/\mathrm{T})$ fewer degrees of freedom than the comparable independent sample design.

It follows from the above discussion that the repeated measurement designs will become efficient only where measures have high reliability or where measures are taken at many ages or many times. Furthermore, it is impossible to separate age differences from practice effects in the cohort-sequential model, and it is impossible to separate time differences from practice effects in the cross-sequential model. The precise meaning of the confounded differences under various assumptions may be obtained from Table 6 by adding practice effect to each of the entries in the age difference column for the cohort-sequential and in the time-lag difference column for the cross-sequential model. These relationships in the case of the repeated measurement designs will hold only if independent sets of samples are used for the cohort- and cross-sequential analyses. Overlapping will, furthermore, add the practice effect to all Td and all Cd terms in the cohort difference column for the cohort-sequential model.[9]

MEASURING THE EFFECT OF PRACTICE

A direct estimate of the effect of practice as a component of developmental change may be obtained from the repeated measurement designs if, and only if, age is not significantly involved in the cohort-sequential design, or if time is not significantly involved in the cross-sequential design. The effect of practice is confounded with the other components of change whenever a more complex set of assumptions is required. This is unfortunate since it is often important to identify the extent to which a behavior has been altered by the measuring operation or to insure that a gain in scores cannot be explained simply by the increased familiarity of the subjects with the measuring instrument. An independent estimate of practice effects may be obtained under all assumptions, however, if the investigator is willing to employ conjointly repeated measurement and independent random sampling designs.

The procedure to be advocated requires that independent samples be drawn which are to be compared with the second or subsequent measurements for the repeated measurement group. The investigator must obtain a random sample from the same cohort which differs from the latter group only in the characteristic that it has been measured one time less than the repeated measurement group. The analysis of variance design for the study of practice effects is shown in Table 9.

TABLE 9

ANALYSIS OF VARIANCE FOR THE CROSS-SECTIONAL STUDY OF PRACTICE EFFECTS

Source of variation	Degrees of freedom
Between ages + cohorts (AC)	$AC-1$
Levels of practice (P)	$P-1$
Age + cohort × practice interaction	$(AC-1)(P-1)$
Error	$N-(AC)(P)$
Total variation	$N-1$

It will be noticed that the proposed design has the usual cross-sectional approach deficiency in that age and cohort differences are confounded and in that the interaction is therefore uninterpretable unless it is known from independent sources that the effect of either age or cohort differences is zero. A far more complex design will be required to remedy this difficulty and is feasible only if an opportunity for at least three measurements exists. In this case, it will also be necessary to obtain a repeated measurement on the random sample first tested on the second occasion. Furthermore, a new random sample must be drawn on the third occasion which differs from both the first and second repeated measurement samples only on the characteristic of having been measured once less than the other groups. Groups may then be split in random halves and appropriate assignment to the different levels of practice made. The analyses for such a design will be similar to the ones given in Tables 3, 4, and 5 except for the addition of a main effect for practice, and the simple and triple interaction between practice and the other components of developmental change. Such a design is quite efficient in the case of the time- and cross-sequential methods. It faces the usual sampling attrition problems in the cohort-sequential approach which for human subjects is, in any event, best limited to short-term studies.

CONCLUSIONS

The results of our analysis suggest that the conventional longitudinal and cross-sectional methods for the study of developmental problems can lead to comparable results only when there are no genetic or cultural changes in response to the variable studied. Such circumstances, however, are rare among the areas of interest to the developmental psychologist. Research strategies are required, therefore, which will permit the unconfounding of the components of developmental change. Consideration of a general model has permitted the development of sequential strategies which permit such unconfounding, and it has furthermore been shown that the conventional methods are simply incomplete special cases of these sequential models. The hazards of the repeated measurement designs have been explored, and alternate methods have been suggested for the study of practice effects.

---

[9] Bell (1953) presents an overlapped design which confounds the information separated in this paper by use of the repeated measurement design for the cohort- and cross-sequential models.

It must now be argued that the cohort-sequential method (and the conventional longitudinal method as a special case thereof) is basically inefficient. Study of Table 6, moreover, will show that the most interesting information available from longitudinal study can equally well be obtained by a combination of the time- and cross-sequential approaches. It is suggested, therefore, that the most efficient design for a developmental study consists of the following scheme:

1. Draw a random sample from each cohort over the age range to be investigated, and measure at Time k (Score A).
2. Get a second measurement on all subjects tested at Time k at Time l (Score B).
3. Draw a new random sample from each cohort in the range tested at Time k plus one cohort below that range, and test at Time l (Score C).

The following comparisons may now be made: (*a*) Cross-sequential model (Scores A and C), (*b*) Time-sequential model (Scores A and C), (*c*) Cross-sequential model with repeated measurements (Scores A and B), (*d*) Cross-sectional model, controlling for effect of practice (Scores B and C).

Almost all questions of interest to the developmental researcher can be handled by this strategy, which should yield interpretable results in practically all empirical situations.[10] Sampling attrition in the repeated measurement part of the study should be fairly limited over a single time interval. Also, if desired, the new random sample obtained at Time l will provide the base for a second repeated measurement study, and any cohort or cohort sequence can, of course, be followed further if a conventional longitudinal study is to be pursued.

It is hoped that this paper has clarified the relations between the different types of basic experimental designs in developmental psychology. A next logical step should be in the direction of reexamining some of the studies in the developmental and attitude survey literature with the hope that puzzling inconsistencies can now be resolved or identified as artifacts of improperly understood experimental designs.[11]

---

[10] The proposed research strategy has been applied to the study of developmental changes in adult cognitive behavior (Schaie & Strother, 1964a; 1964b).

[11] Examples of studies containing data which could readily be examined via the proposed models may be found in Shuttleworth (1939) and in Hyman and Sheatsley (1964).

REFERENCES

Bell, R. Q. Convergence: An accelerated longitudinal approach. *Child Development*, 1953, **24**, 145–152.

Horst, P. Multivariate models for evaluating change. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: Univer. of Wisconsin Press, 1963.

Hyman, H. H., & Sheatsley, P. B. Attitudes toward desegregation. *Scientific American,* 1964, **211**, 16–23.

Kessen, W. Research design in the study of developmental problems. In P. Mussen (Ed.), *Handbook of research methods in child development.* New York: Wiley, 1960.

Kuhlen, R. G. Social change: A neglected factor in psychological studies of the life span. *School and Society*, 1940, **52**, 14–16.

Schaie, K. W. Cross-sectional methods in the study of psychological aspects of aging. *Journal of Gerontology*, 1959, **14**, 208–215.

Schaie, K. W., & Strother, C. R. A cross-sequential study of age changes in cognitive behavior. Paper read at Midwestern Psychological Association, St. Louis, April–May 1964. (a)

Schaie, K. W., & Strother, C. R. The effect of time and cohort differences on the interpretation of age changes in cognitive behavior. *American Psychologist*, 1964, **19**, (Abstract) 546. (b)

Shuttleworth, F. K. The physical and mental growth of girls and boys age six to nineteen in relation to age at maximum growth. *Monographs of the Society for Research in Child Development*, 1939, No. 4.

(Received September 30, 1964)

Psychological Bulletin
1965, Vol. 64, No. 2, 108–112

A MODIFICATION OF THURSTONE'S LAW OF COMPARATIVE JUDGMENT TO ACCOMMODATE A JUDGMENT CATEGORY OF "EQUAL" OR "NO DIFFERENCE"[1]

MARSHALL G. GREENBERG[2]

University of Minnesota

A model is proposed which modifies the Thurstone law of comparative judgment extending it to the treatment of pair comparison data in which judges are permitted a judgment category of "equal" or "no difference." Based upon a rational description of the perceptual process the model introduces a threshold parameter and apportions the "equal" judgments between the 2 stimuli to estimate the actual proportions perceiving one greater than the other on some underlying attribute. This apportionment is carried out within the spirit of the Thurstone model and is shown to be closely approximated by splitting the "no difference" responses equally between the 2 stimuli being compared. Problems of estimation and alternative solutions are discussed as well.

In 1927 a pair of classic papers published by Thurstone (1927a, 1927b) introduced his law of comparative judgment thereby giving birth to a new era in psychological scaling. This model provided a means for analyzing not only comparisons of physical intensities, but qualitative judgments as well, and led to applications in the measurement of attitudes and other subjective attributes. In brief, the law of comparative judgment provides a means for transforming the pair comparison data obtained from the $k(k-1)/2$ pairs generated by k stimuli into a set of k scale values for these stimuli along some underlying psychological or psychophysical dimension. Thurstone specified (1927a) that the method "applies fundamentally to the judgments of a *single observer* who compares a series of stimuli by the method of paired comparison when no 'equal' judgments are allowed [p. 276]." He then proceeded to modify the model to permit replications of judgments across as well as within individuals. The present paper introduces a modification which permits judgments of "equal" or "no difference" while retaining the spirit of Thurstone's approach.

THURSTONE'S MODEL

No attempt will be made to describe Thurstone's law of comparative judgment in great detail since it has been presented and summarized in many places elsewhere (Coombs, 1964; Guilford, 1954; Thurstone, 1927a, 1927b; Torgerson, 1958). The present discussion simply attempts to capture the flavor of the model while laying the groundwork for the modification to follow.

The model assumes that when a pair of stimuli is being judged with respect to some attribute, each stimulus may be represented by a distribution over the continuum as in Figure 1. The use of distributions rather than point representations reflects the notion that there exist momentary fluctuations in the individual's perceptions of the stimuli. Designating the entire set of stimuli by $S = \{S_1, S_2, \cdots, S_i, \cdots, S_k\}$ the distribution for any stimulus S_i is assumed to be normal with mean $\bar{S}_i$ and variance σ_i^2. When an individual is presented with two stimuli and is asked to judge which has more of some attribute it is assumed that he draws a random sample from each of the two stimulus distributions and chooses the stimulus whose sample observation is the greater. Since the decision is dependent upon the difference between sample observations, the model may be formulated in terms of the distribution of sample differences as shown in Figure 2.

Given the individual distributions for S_i and S_j, the difference distribution is known to be normally distributed with mean $\bar{S}_j - \bar{S}_i$ and variance $\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j$. The shaded area in Figure 2 corresponds to the proportion of

---

[1] This work was supported in part by Grant GB-2345 from the National Science Foundation.

[2] Now at The Procter & Gamble Company, Cincinnati, Ohio.

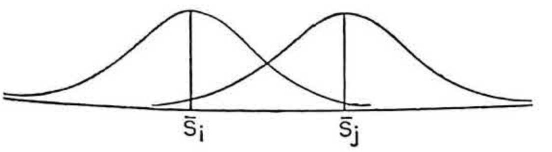


FIG. 1. Representations of two stimuli along an attribute.

times S_i is perceived as greater than S_j and the complementary proportion is the proportion of times S_j is judged to be greater than S_i. Using empirical data to estimate these proportions, one attempts to determine the scale values $\bar{S}_1, \bar{S}_2, \cdots, \bar{S}_i, \cdots, \bar{S}_k$ of the k stimuli by solving a set of simultaneous observational equations of the form:

$$\bar{S}_i - \bar{S}_j = z_{ij}(\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j)^{1/2}, \quad [1]$$

where z_{ij} is the normal deviate corresponding to the proportion of judgments of S_i greater than S_j. Because there are many more unknowns than observational equations, special cases need to be established in order to determine a solution.

Thurstone's Case V is the one which is most frequently employed and makes the following set of simplifying assumptions:

1. Intraindividual replications can be replaced by interindividual replications.
2. $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_i^2 = \cdots = \sigma_k^2 = \sigma^2$
3.[3] $r_{12} = r_{13} = \cdots = r_{jk} = r$

Thus Equation 1 becomes

$$\bar{S}_i - \bar{S}_j = z_{ij}\sigma[2(1-r)]^{1/2}. \quad [2]$$

By letting $\sigma[2(1-r)]^{1/2}$ be the unit of measurement and setting it equal to one, we obtain

$$\bar{S}_i - \bar{S}_j = z_{ij} \quad [3]$$

as the observational equation.

Methods for solving a set of such equations are described by Thurstone (1927a), Guilford (1954), and Torgerson (1958). Mosteller (1951) has shown that these procedures yield a least-squares solution for estimates of the scale values. Mosteller also discusses the solution for cases in which the standard deviations are unequal and presents a chi-square goodness-of-fit test for the case where the sigmas are equal.

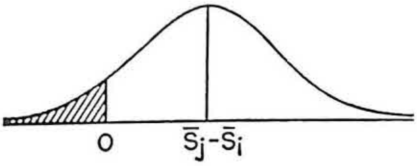


FIG. 2. Distribution of sample differences for S_i and S_j.

A MODIFICATION OF THURSTONE'S MODEL

Thurstone's model assumes that whenever the sample observation drawn from S_j is greater than that drawn from S_i the individual chooses S_j as having more of the relevant attribute. Since the distributions involved are continuous the probability of equal sample observations is zero, and a binary response will always occur. The proposed modification introduces a parameter k_{ij} analogous to a threshold value representing the smallest difference between sample observations that can be detected by the individual. When the sample observations differ by an amount greater than k_{ij} the stimulus generating the larger sample observation is chosen. When they differ by less than k_{ij} a judgment of equality is offered. Letting s_i represent a sample observation from S_i,

$$s_i - s_j > k_{ij} \Rightarrow S_i \text{ chosen}, \quad [4]$$

$$s_i - s_j < -k_{ij} \Rightarrow S_j \text{ chosen}, \quad [5]$$

$$-k_{ij} \le (s_i - s_j) \le k_{ij} \Rightarrow S_i \text{ and } S_j \text{ judged equal}. \quad [6]$$

The present model segments the distribution of differences as shown in Figure 3. The shaded area corresponds to the proportion of judgments of "equal," "same," or "no difference,"

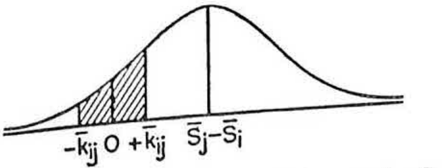


FIG. 3. Distribution of sample differences for S_i and S_j with threshold parameter $\bar{k}_{ij}$.

---

[3] Thurstone actually assumed that all the intercorrelations were zero. Mosteller (1951) pointed out that this assumption was never used and that in fact it need only be assumed that all the intercorrelations are equal.

while the other two sections represent the proportion of responses of $S_i > S_j$ and $S_j > S_i$.

At this point we may further introduce the assumption that k_{ij} is a normally distributed variable with mean $\bar{k}_{ij}$ and variance $\sigma^2_{k_{ij}} = \sigma^2_k$ for all i and j. Then if the values of k are assumed to be independent of the $S_i - S_j$ joint distributions the effect is simply to increment the variance of the difference distributions by σ^2_k. This will alter the unit of measurement but leave the solution otherwise unchanged.

Once again we need to solve equations of the form

$$\bar{S}_i - \bar{S}_j = z'_{ij} \qquad [7]$$

where z'_{ij} is the standardized distance between $(\bar{S}_i - \bar{S}_j)$ and 0. The relevant normal deviate, however, no longer corresponds to one of the observed proportions, but is midway between their normal transformations.

Let p_{ij} and p_{ji} be the proportions of judgments $S_i > S_j$ and $S_j > S_i$, respectively. We lose no generality by assuming $p_{ij} < p_{ji}$. Then

$$\bar{S}_j - \bar{S}_i - \bar{k}_{ij} = z_{ji} \qquad [8]$$

and

$$\bar{S}_j - \bar{S}_i + \bar{k}_{ij} = - z_{ij}, \qquad [9]$$

where z_{ij} and z_{ji} are the normal deviates corresponding to p_{ij} and p_{ji}, respectively. Summing Equations 8 and 9 yields

$$\bar{S}_j - \bar{S}_i = 1/2(z_{ji} - z_{ij}) = z'_{ji}. \qquad [10]$$

The z' values obtained from Equation 10 may now be incorporated into the usual solutions for the law of comparative judgment model to recover the set of scale values $\bar{S}_1, \bar{S}_2, \cdots, \bar{S}_i, \cdots, \bar{S}_k$.

DISCUSSION OF THE MODEL[4]

The proposed modification of the Thurstone model contributes to its application in essentially two ways: (*a*) It permits the collection of pair comparison data in which individuals have been offered the option of judging stimuli to be equal. (*b*) It apportions judgments of equality to the two stimuli in a pair comparison in a manner consistent with the spirit of Thurstone's model. In fact the original model becomes the special case of the modified version in which $\bar{k}_{ij} = 0$ for all i and j.

A difficulty inherent in the usual solutions to the standard Thurstone model arises in connection with this modified version as well. There are two sets of parameters to be estimated in the present model, a set of $\bar{k}_{ij}$ values and a set of interstimulus distances of the form $\bar{S}_j - \bar{S}_i$. The interstimulus distances are estimated from Equation 10. The expression for estimating the discrimination threshold is obtained by subtracting Equation 9 from 8 which yields

$$\bar{k}_{ij} = - 1/2(z_{ij} + z_{ji}). \qquad [11]$$

Denoting the asymptotic covariance of these estimates by $\gamma_{\bar{k}_{ij}, \bar{S}_j - S_i}$, it follows from Equations 10 and 11 that

$$\gamma_{\bar{k}_{ij}, \bar{S}_j - S_i} = 1/4(s^2_{z_{ij}} - s^2_{z_{ji}})$$
$$= 1/4\left[\frac{p_{ij}(1 - p_{ij})}{Ny^2_{z_{ij}}} - \frac{p_{ji}(1 - p_{ji})}{Ny^2_{z_{ji}}}\right], \qquad [12]$$

where $s^2_{z_{ij}}$ is the sampling variance of z_{ij} as estimated from the sample proportion, p_{ij}, $y_{z_{ij}}$ is the ordinate of the normal curve at the point of truncation z_{ij}, and N is the total number of judgments. To the extent that the sampling errors of z_{ij} and z_{ji} differ there is a covariance between the estimates of the two parameters involved. Thus, a least-squares solution to the problem as it stands is quite difficult. There are at least two ways of resolving the difficulty presented by this covariance.

The approach taken in the present model, suggesting the use of Equations 10 and 11, ignores the covariance problem. In most applications in which the model would be employed, the proportion of no-decision votes will be low, yielding small values for the $\bar{k}_{ij}$'s. Consequently differences in the sampling errors of z_{ij} and z_{ji} will be minimal. This approach essentially parallels that of the standard Thurstone procedure as described by Mosteller (1951). Mosteller's least-squares solution also fails to take into consideration differences in the variabilities of the z_{ij} values used to estimate scale separations. A true least-squares solution would assign weights to the estimates in proportion to their sampling errors, but Mosteller pointed out that such methods would be "unmercifully extensive." An additional justification for ignoring the correlated estimates is the fact that the primary interest lies in estimating the scale separations while the estimation of the $\bar{k}_{ij}$ values is tangential to this more basic problem. One can apply Equation 10, in which the value of $\bar{k}_{ij}$ has dropped out, to obtain an estimate of z'_{ij} and go on to employ the usual Thurstone analysis without ever considering Equation 11.

---

[4] After this paper had been submitted for publication the author discovered that Glenn and David (1960) had independently conceived and developed the same modification of the Thurstone model, although their analytic procedure differed from that proposed here. This portion of the present paper has undergone considerable revision as a result of the ideas presented in their paper.

A second approach is that taken by Glenn and David (1960). In order to avoid the problem of correlated estimates they approximated the normal difference distribution by a sine function and employed an inverse sine transformation to scale the response proportions. Glenn and David showed that for large samples the covariance between estimates of scale distances and the discrimination thresholds approaches zero in such a model. Still faced with slightly nonhomogeneous variances among estimates of scale separations and the discrimination thresholds, the authors presented both weighted and unweighted solutions. Their unweighted solution is essentially the one presented in this paper, but employing an inverse sine transformation rather than a normal transformation on the sample proportions. Their weighted solution uses the results of the unweighted analysis as the first step of a rather extensive iterative procedure which would be impractical without the aid of an electronic computer. Glenn and David supplied a numerical example in which they obtained scale values for secretaries' preferences among five brands of carbon paper. When Equation 10 of the present model was applied to their data and the z'_{ij} values analyzed by the standard Thurstone procedure, the resulting scale values produced Pearson product-moment correlations of .999 with both their weighted and unweighted solutions.

The utility of the model for any particular application depends upon the assumptions one is willing to make and the knowledge one has regarding the properties of the parameter k. For studies in which the p_{ij} values are based upon intraindividual replications, the application of Equation 10 and the subsequent solution is quite straightforward. To replace intraindividual replications with interindividual replications it is necessary to assume that all of the judges have approximately the same capacities for discrimination. Even if the values of $\bar{k}_{ij}$ vary across individuals the model may satisfactorily represent the data by yielding values of $\bar{k}_{ij}$ which are averages of the individual threshold values. If each individual makes a large number of judgments it is possible to evaluate the assumption that the different judges have the same discrimination threshold. The procedure involves comparing the abilities of the solutions to reproduce the obtained proportions when a common $\bar{k}_{ij}$ is assumed and when it is not. A chi-square goodness-of-fit test to a trinomial distribution is required and is described by Glenn and David (1960).

A common problem in political polling, market research, and other settings in which binary responses are requested, but judgments of "equal" or "no difference" occur, is how to allocate the "undecided" votes between the two stimuli being compared. There are two frequently employed methods of resolving this problem. The first is to split the "undecided" votes equally between the two stimuli; the second is to allocate these responses proportionately by dividing each of the choice proportions by their sum so that the resulting values add to one. The present model produces a result somewhere between these two methods. To the extent that the integral of the normal difference distribution between 0 and $-\bar{k}_{ij}$ is equal to that between 0 and $+\bar{k}_{ij}$ the "equal" votes will tend to be split 50–50 between S_i and S_j. These integrals or areas will tend to be equal whenever either of two conditions holds: (*a*) The proportion of S_i judgments approaches the proportion of S_j judgments. (*b*) The proportion of "equal" judgments is very small (i.e., $\bar{k}_{ij}$ is small). It is interesting to note that Thurstone (1927a, p. 277) discussed the discrepancies between these two areas for very small deviations from 0 on the difference distribution. Table 1 illustrates the results, for a number of hypothetical splits, of the three methods of allocating the "equal" votes. Clearly the modified Thurstone model is almost as conservative in its allocation of the "equal"

TABLE 1

RESULTS OF APPORTIONING "EQUAL" VOTES BY THREE DIFFERENT METHODS

Data			Proportional split		50–50 split		Modified Thurstone model	
S_i %	S_j %	Equal %	S_i %	S_j %	S_i %	S_j %	S_i %	S_j %
40	40	20	50.0	50.0	50.0	50.0	50.0	50.0
50	30	20	62.5	37.5	60.0	40.0	60.3	39.7
55	25	20	68.8	31.2	65.0	35.0	65.5	34.5
60	20	20	75.0	25.0	70.0	30.0	70.8	29.2
60	10	30	85.7	14.3	75.0	25.0	77.9	22.1
70	20	10	77.7	22.3	75.0	25.0	75.2	24.8
70	10	20	87.5	12.5	80.0	20.0	81.6	18.4

responses as is the technique of distributing them equally. Conservative as used here refers to the probability of a Type I error associated with a test of the difference between the resulting two percentages. For most applications where the present model is relevant it would appear quite satisfactory to permit respondents to employ an equality category and to apportion such responses equally between the alternative stimuli followed by the standard Case V solution, although the use of Equation 10 requires little extra effort.

For any given area of investigation one might test the validity of the present model by experimentally manipulating the threshold parameter k. This could be accomplished by employing different degrees of encouragement to avoid using the "no difference" category in the instructions to judges or by imposing varying time limits on the choice such that if the limit is exceeded the response is recorded as an inability to discriminate. Applying the model to such data in which k has been varied should yield linearly related sets of scale values for differing values of k.

REFERENCES

COOMBS, C. H. *A theory of data.* New York: Wiley, 1964.

GLENN, W. A., & DAVID, H. A. Ties in paired-comparison experiments using a modified Thurstone-Mosteller model. *Biometrics*, 1960, **16**, 86–109.

GUILFORD, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.

MOSTELLER, F. Remarks on the method of paired comparisons. Parts I, II, and III. *Psychometrika*, 1951, **16**, 3–9, 203–218.

THURSTONE, L. L. A law of comparative judgment. *Psychological Review*, 1927, 34, 273–286. (a)

THURSTONE, L. L. Psychophysical analysis. *American Journal of Psychology*, 1927, 38, 368–389. (b)

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

(Received October 1, 1964)

Psychological Bulletin
1965, Vol. 64, No. 2, 113–123

NEURAL SYSTEMS OF RESPONSE RELEASE AND CONTROL

LAUREN K. GERBRANDT [1]

Claremont Graduate School

A descriptive model of control and release functions of behavior was derived from recent brain research literature on this topic. The model allowed comparisons to be made between several common behavioral tasks as well as within the acquisition to retention phases on each of these tasks. Electrophysiological, neurological, and psychopharmacological effects known in the literature were seen to be ordered by the model which assumed specified mutually inhibitory brain systems to underlie the behavioral functions of release and control of responses.

The importance of the situational demands of "the control and release of behavior" has long been recognized by Pavlov (1928) in his studies of neural inhibition and excitation phenomena. Pavlov inferred these mutually antagonistic neural events of excitation and inhibition from the behavioral indices of conditioned reflexes and the conditioned or unconditioned inhibitions of these reflexes. The present paper will expand and elaborate on the list of behavioral indices and will examine endogenous and exogenous changes in specified mutually inhibitory brain substrates as they relate to these examples of behavior release and control. The descriptive model of such behavioral indices will first be formed from the recent literature of this country related to brain research on behavioral release and control systems. This model will then be related systematically to neurological, electrophysiological, and drug research.

RELEASE AND CONTROL

Suggestions of behavioral release and control factors may be seen in the discussions of several recent authors in this country. The earliest of these is the suggestion of the concept of "act inhibition" to explain the behavioral deficits in delayed response performance after frontal lobe lesions (Stanley & Jaynes, 1949). Other authors (McCleary, 1961; Schwartzbaum, Kellicutt, Spieth, & Thompson, 1964) have used the similar constructs of "response inhibition and facilitation" to explain the behavioral deficits obtained on types of avoidance learning. Thomas and Slotnick (1963) and Vanderwolf (1964) have used the related constructs of "release and inhibition of a freezing mechanism" to explain their effects on avoidance performance. The suggestion that lesions may "release or control motility mechanisms" (Maher & McIntyre, 1960) differs by specifically denoting the classes of behavior either released or controlled. Less similar are the constructs used by various authors attempting to explain lesion drug, or electrophysiological effects on these same behavioral tasks. MacAdam (1962), and Chalmers and Erickson (1964) have referred to "tonic and phasic orienting response mechanisms" as being released or controlled, thereby interfering with the performances of some but not other tasks. Gerbrandt (1964) has used the similar constructs of "control and release of attentive shifting mechanisms" to explain various behavioral effects produced by lesions.

Specific examples of behavioral release and control may be seen in the descriptions of the tasks used by McCleary (1961) and others (Ellen, Wilson, & Powell, 1964; Webster & Voneida, 1964). To obtain evidence of response inhibition, McCleary first trained lesioned cats to approach a food station; he then punished the cats for such approaches. Since the animals were required to control their approaches, the punishment phase was used as the critical test for inhibition. To test for response facilitation on the other hand, McCleary trained his cats to release an avoidance response of running upon a signal presen-

---

[1] This paper was prepared during the tenure under Research Fellowship MH-21,492 from the National Institutes of Mental Health.

tation. He found that these two tasks dissociated the effects of two different limbic system lesions. Ellen, Wilson, and Powell (1964) described the effects they obtained from septal lesions on timing behavior as "the result of impairment in the giving up [control] of old responses rather than in the learning of new response patterns." Similarly, Webster and Voneida (1964) described the effects of hippocampal lesions on reversal learning as a deficit in the "altering [control] of a previously stabilized habit." These authors could well have been describing the passive avoidance task of McCleary (1961) or the conditioned inhibition task of Pavlov (1928). From each of these descriptions a composite has been formed to circumscribe the realm of behaviors to be examined in this paper. It is stated as follows:

Discrete brain systems, mutually inhibitory in their effects, function on the one hand to release highly stabilized responses, and on the other hand to allow increases in the stability of elicitation of a response by control over competing responses of higher stability.

By discrete and mutually inhibitory systems of responding, it is meant that the operation of either identifiable neural system is inhibitory on the other, and vice versa. As used in this paper, a behavior is stabilized relative to its susceptibility to interference by "external inhibition." A clue to which behaviors may then fall in the range of "unstabilized" or "stabilized" may be seen in the following quote from Pavlov (1928):

> The more experiments we make with conditioned reflexes . . . the more we have proofs showing that the process of internal inhibition is more *labile* (unstable) than the process of conditioned excitation, i.e., under the influence of extraneous stimuli the process of internal inhibition is suppressed more easily and more quickly than the process of conditioned excitation [p. 239].

Common examples of stabilized responses may be seen then in the performances of conditioned responses, simultaneous discriminations, and conditioned avoidance responses. Common examples of unstabilized behaviors are passive avoidance responding, reversal learning, extinction, and conditioned inhibition responding. Although stability may be generally said to be greater in the former behaviors and less in the latter behaviors, an important stipulation on the degree of stability in these behaviors is that all responses used to infer internal inhibition eventually phase out into excitation, that is, they become stabilized. This warning is based on the finding by Winnick and Hunt (1951) that a buzzer which inhibited extinction was only effective in the early trials of extinction training. Winnick and Hunt also found that a conditioned running response was externally inhibited to a greater extent in early trials than in later trials of acquisition. It might then also be expected that the very early phase of acquisition of conditioned responses would be comparable to other more representative unstabilized behaviors (passive avoidance, etc.) in the functioning of the proposed brain systems of control. The electrophysiological, lesion, and drug effects will now be examined as they relate to the above descriptive model in order to investigate the discreteness of the hypothesized underlying brain systems.

Electrophysiological Recordings

Electrophysiological recording is especially useful for following the activities of many brain substrates simultaneously and temporally during the progress of learning since it allows behaviors to develop without interference from treatments. In that the early phase of acquisition is representative of increases in the stability of behaviors (and concurrent control of competing responses), while the late phase is characteristic of stabilized responses, a "phasing out" of some neural activities should be expected during the course of learning while a "phasing in" of others should take place if discrete brain events underlie these different behavioral functions. In a previous examination of the neurological literature (Gerbrandt, 1964), it was suggested that the first phase might be identified with neural operations in the second and third rhinencephalic systems of Pribram and Kruger (1954), while the late phase functions might be characteristic of the nonspecific projection system (Nauta & Whitlock, 1954). Consistently, Morrell (1957) has found hippocampal single unit firing to conditioned flicker following during the early stages of training, where units of the thalamic nucleus ventralis anterioralis (nonspecific) were then inactive. As training progressed, hippocampal units were found to drop out while the thalamic units increased to a plateau with the stabilized flicker following. John and Killam (1959) trained an active avoidance response to a flicker conditioned stimulus (CS) while recording electroencephalographic (EEG) activity in the hippocampus. They found hippocampal following of the flicker CS during early but not during late training trials. Since the temporal superimposition of a negative CS into the training trials constitutes an unstabilized behavioral situation (the negative conditioned response is in competition with a trained positive CR which must be controlled), it is interesting that John and Killam also found that hippocampal following to the negative CS continued after the dropout to the positive CS. Grastyan, Lissak, Madarasz, and Donhoffer (1959) tested for hippocampal EEG activities during the acquisition of both classical and defensive conditioning. In early trials of either task, where the animals were excited by the CS but had not yet reliably demonstrated task performance, synchronous slow wave activity (theta rhythm) was recorded in the hippocampus. Upon stabilization of conditioned responding, however, the slow wave activity was phased out into "desynchronization" or fast activity. Grastyan et al. also found that differential inhibition and extinction procedures resulted in a temporary return of the theta rhythm accompanying the positive and negative CS and, as in the study by John and Killam, with continued differential inhibition training, the theta rhythm first disappeared to the positive CS. The classical aversive conditioning results of MacAdam (1962) are similar to those of Grastyan et al. and support at a macrorecording level the single unit findings of Morrell (1957) for the hippocampal and thalamic activities during the course of training.

These studies indicate an overwhelming support for the notion that the hippocampus is involved in the phasing out of a control over competing responses, and that a consequent phasing in of release systems is occurring in other brain areas. Unfortunately, the results of Adey, Dunlop, and Hendrix (1960) are in seeming contradiction with the above reported studies. Adey et al. recorded EEG activities from several regions of the hippocampus and found that the greatest synchrony of slow wave activity occurred during the late phase of stabilized approach of performance, and during the actual motor act of food approach within a given trial. These results have since been corroborated by Bremner (1964), and by Vanderwolf and Heron (1964). Grastyan et al. (1959) reported EEG desynchrony during this phase and identified a relation between orienting responses and slow waves from a photographic analysis of the approaches. Adey et al. reported no relation of slow wave activity to orienting responses from their photographic analysis. Several possibilities exist for making these data consistent. Grastyan et al. reported cases where one electrode in an animal would be mostly synchronous where yet another was desynchronous, namely, during an actual food approach. Although it would seem that both a Grastyan et al. and an Adey et al. preparation have occurred here in the same animal, in both of these studies it is reported that the slow wave activities are best recorded in the apical dendrites of the pyramidal cells of the hippocampus. A second possibility for consistency may be seen in other details of the Adey et al. findings. During early phases of learning, these authors found theta activity to proceed temporally from septohypothalamic inputs to hippocampal areas CA_4, to CA_1, and to exit out the subiculum into the entorhinal areas. During later phases of training, the temporal course of wave appearances was reversed. If the Grastyan et al. recordings of slow waves could be identified with the early phase relations of Adey et al., then it might be claimed that the late phase slow waves were present but masked in the records of the former investigators. Identification of slow waves with orienting responses then might be properly made only during the early phase activities. Finally, differences in task parameters might be experimentally manipulated for clues to the discrepancy. Grastyan et al. have found hippocampal desynchrony to be most reliably found in late trials and synchrony in early trials of daily sessions. Also, it is not inconceivable that the unitary response (Grastyan et al.) and choice response (Adey et al.) differences in task parameter may produce functional dichotomies (Bitterman, 1962).

In spite of the fact that these differences have not yet been brought into accord, some tentative hypotheses might be made in light of other evidence. Petsche and Stumpf (1960) have identified the pacemaker of hippocampal slow wave activities as lying in the medial septal nuclei. Septal lesions are known to irreversibly eliminate hippocampal slow waves (Green & Arduini, 1954). Septal lesions are also known to produce specific deficits on passive avoidance (McCleary, 1961), on reversal learning (Thompson & Langer, 1963), and on the pacing of bar-pressing responses (Ellen, Wilson, & Powell, 1964). Septal lesions are known to facilitate conditioned avoidance responding (Krieckhaus, Simmons, Thomas, & Kenyon, 1964; McCleary, 1961). Induction of theta activity with injections of physostigmine (Mayer & Stumpf, 1958), or medial septal implants of cholinergic stimulants, will block the *performance* (late phase) of a conditioned avoidance response (CAR) (Chalmers & Erickson, 1964; Grossman, 1964). Scopalamine and atropinic drugs, which block the EEG effects from injection of cholinergic stimulants (Green, Maxwell, Schindler, & Stumpf, 1960), block only the *acquisition* (early phase) and not the performance of a CAR (Meyers, Roberts, Riciputi, & Domino, 1964). Chalmers and Erickson (1964) also reported that atropine blocks the physostigmine-produced interference with the performance of a CAR. It would appear that the hippocampal theta rhythm, having its source in the septal nuclei and proceeding through pyramidal cells of the hippocampus to the entorhinal area, is necessary for control of competing responses while conditioned behavior is stabilizing. It would appear also that the hippocampus mediates a slow wave activity identified with entorhinal, rather than septohypothalamic, inputs. This slow wave activity, however, is characteristic of late learning phase functioning. These hypotheses will be reexamined in light of lesion and drug studies in the following sections.

LESION EFFECTS

The notion that discrete and mutually inhibitory brain systems operate during the control and release of stabilized responses suggests that during some task performances and not others the brain will have identifiable bias characteristics. The endogenous biases were seen in many studies discussed in the previous section. In this section the concern will be with the experimental biases produced by various lesions. Since such biases are relevant to discrete systems, rather than centers, of neural activity, knowledge of the projections and inputs to lesioned nuclei will be more crucial in interpreting lesion effects than the knowledge of which nuclei are damaged per se.

From the discrepancies in electrophysiological evidence previously discussed (Adey et al., 1960; Grastyan et al., 1959), it might be expected that contradictory findings will occur in lesion studies on the hippocampus. It is not too surprizing then that, while Isaacson, Douglas, and Moore (1961) found no hippocampal lesion effect on active avoidance, Hunt, Diamond, Moore, and Harvey (1957) did find such a *performance* (but not acquisition) effect. Buresova, Bures, Fifkova, Vinogradova, and Weiss (1962), on the other hand, reported an impairment on an active avoidance *acquisition*. Utilizing the same type of passive avoidance task, Kimble (1965 unpublished) reported no hippocampal lesion effect while Teitelbaum and Milner (1963) found large performance losses. Using a passive avoidance task where a motivated approach is punished, Kaada, Rasmussen, and Kveim (1962), and Kveim, Setekleiv, and Kaada (1964) found no hippocampal lesion deficits while such deficits were reported by Isaacson and Wickelgren (1962), and by Kimble (1963). Results on hippocampal lesions and reversal learning, however, are consistent in showing reversal losses without affecting performance of the task to be reversed (Mahut & Cordeau, 1963; Teitelbaum, 1964; Thompson & Langer, 1963; Webster & Voneida, 1964). Although contradictions are expected while using a gross technique such as lesioning, there is very little clue within these studies for purposes of resolution. In other studies, however, Kimura (1958) reported dorsal but not ventral hippocampal lesions to disrupt passive avoidance responding. Differential projections and activities from the dorsal and ventral hippocampi have been reported (Elul, 1964). Rosvold and Szwarcbart (1964) also reported functional separation in different lesions, in that anterior hippocampal lesions produced greater delayed alternation deficits than posterior lesions, while lesions within hippocampal regions CA_1 and CA_4 produced greater deficits than in regions CA_2 and CA_3. The latter finding may be related to the report by Adey et al. (1960) that regions CA_1 and CA_4 are the respective inputs for late- and early-phase relations. Although region CA_4 should have been the more relevant site producing losses on delayed alternation, from a vertical lesioning approach an attempt at lesioning only the CA_1 region may frequently involve region CA_4 as well. It would be more revealing in this regard if animals were lesioned across the fiber tracts entering region CA_1 from the entorhinal cortex and were tested for acquisition and retention of a CAR task. Since entorhinal lesions apparently do not abolish hippocampal slow wave activities (Adey, Sunderland, & Dunlop, 1957), while septal lesions do (Green & Arduini, 1954; Petsche & Stumpf, 1960), entorhinal damage should not affect acquisition of the CAR. Since the late phase functioning is related to entorhinal inputs, such damage should affect losses on retention of the CAR.

While damage to the medial septal nuclei does lead to abolition of hippocampal slow wave activities, clear predictions of such lesion effects on behavioral tasks cannot be made since, as previously noted, Adey et al. (1960) reported slow wave activity during stabilized approaches only, while Grastyan et al. (1959) noted slow waves to occur only during the conditioning phase previous to stabilization of performance. Either there are different pacemakers for the slow wave activities identified in the two phases of responding, in which case medial septal lesions will produce losses only on acquisition and other situations of competing response control, or it is a pacemaker for both phases and is differentiated either by activating the hippocampus from both indirect (entorhinal) and direct (fornix) septal inputs, or by some chemical or electrical specificity. Only very tentative conclusions bearing on these alternatives may be drawn from the literature reported thus far. Septal lesions have been found, as previously reported, to affect passive avoidance (McCleary, 1961), reversal learning (Thompson & Langer, 1963), extinction (Schwartzbaum, Kellicutt, Spieth, & Thompson, 1964), differential inhibition (Schwartzbaum et al., 1964), and conditioned inhibition (Brutkowski & Mempel, 1961). Many investigators also reported facilitated performances as a result of septal lesioning of animals performing a CAR (King, 1958; Krieckhaus, Simmons, Thomas, & Kenyon, 1964; McCleary, 1961). Moore (1964), however, has reported septal lesions to produce deficits in the retention of a CAR. The results of Vanderwolf (1964) and of Kenyon (1962) were not in accord with his findings since they reported decrements in the *acquisition* but not the *retention* of a CAR. It must not be concluded, however, that both control and release functions are therefore mediated by the same septal brain cells; none of these studies, other than McCleary (1961), tested for both passive and active avoidance performances in the same animals. McCleary (McCleary, Jones, & Ursin, 1965) has again followed up these septal lesion study discrepancies in such a fashion. In this study, McCleary deliberately placed septal lesions in anterior and posterior groups.[2] He found anteriorly placed septal lesions, especially in the medial septal nucleus, to produce deficits on the passive but not on the active avoidance task. Posteriorly placed lesions, however, produced no losses on the passive but did produce deficits on the active avoidance task. These findings are quite interesting in light of the degeneration studies of Powell (1963) which showed anterior septal lesioned sites to be differentially projecting to the hippocampus over fornix fibers, to the anterior cingulate area over the fornix longus, and to the tegmentum via the medial forebrain bundle. Correspondingly, lesions of the precommissural fibers produce deficits on reversal learning (Thompson & Langer, 1963) and passive avoidance responding (Kaada, Rasmussen, & Kveim, 1962); lesions of the anterior cingulate cortex produce losses on a passive but not an active avoidance task (McCleary, 1961) and on a conditioned inhibition task (Brutkowski, 1957); and lesions of the medial forebrain bundle produce losses on passive avoidance responding (Kaada, Rasmussen, & Kveim, 1962) and reversal learning (Thompson & Langer, 1963). Many of these brain areas are known to induce septal-hippocampal slow wave activity upon electrical stimulation (Green & Arduini, 1954). Powell (1963) also found that posterior septal lesions produced differential degeneration in the thalamic anteroventral, reticular, and habenular nuclei. Consistently, lesions of the anteroventral thalamic (Thomas & Slotnick, 1962), as well as the habenular (Thompson, 1963), nuclei have been reported to produce deficits on a CAR performance. Gerbrandt (1965) reported no effect on learning a passive avoidance task in animals lesioned in habenular, as well as dorsomedial, thalamic brain areas. When the differential projections of lesioned areas within the septal nuclei are taken into account, then, findings suggest anatomical separations for the control and release behavioral functions. Continued investigation is needed, however, to determine the relation between EEG activities as they originate from these areas and the functions they serve.

---

[2] R. A. McCleary, personal communication, 1964.

The remainder of the lesion studies indicating production of a neural bias in control and release systems have largely been reviewed in a previous paper (Gerbrandt, 1964). To reiterate, these studies indicated a circuit, roughly comparable to that of Papez (1937), functioning to control behavior competing with a response to be stabilized. The nonspecific projection system, on the other hand, was implicated in the release of stabilized responses. Aside from those lesion studies already reported in this paper, several relevant studies have appeared since the preparation of the previous review. The expected findings that mammillary and mammillothalamic tract lesions would not interfere with shuttlebox avoidance responding, have been supported by the investigations of Dahl, Ingram, and Knott (1962), and of Ploog and MacLean (1963). Again, the opposite results have been found by Thomas, Fry, Fry, Slotnick, and Krieckhaus (1963), and by Krieckhaus (1964); these authors found losses on the retention but not the acquisition of shuttlebox and one-way active avoidance responding. A knowledge of fiber degenerations seems also to be crucial in interpreting these results. Krieckhaus (1964) noted that the fibers running from the lateral mammillary bodies to the anterodorsal thalamic nuclei may be responsible for these effects. Consistently, the anterodorsal thalamic nucleus sends a heavy projection to the posterior cingulate cortex (Locke, Angevine, & Yakovlev, 1964), which has been reported to produce losses in active avoidance responding (Thomas & Slotnick, 1963). In the rat, where strong differential projections from the septum, via the mammillary bodies, to the anteromedial thalamic nuclei have been described (Powell, 1963), the mammillary body and tract and the anteromedial thalamic nuclei lesions produce reversal (Thompson & Langer, 1963) and passive avoidance learning (Gerbrandt, 1965) deficits. Future investigations are greatly needed which will determine whether such losses may be produced in the same animals where no losses are reported on other tasks of opposed function (Dahl, Ingram, & Knott, 1962; Ploog & MacLean, 1963). Until then, these hypotheses remain ad hoc.

The assumption of interactions between brain systems, rather than between centers, has been illuminated other than by contradictory data reported for the same nucleus. A mutual inhibition between two systems implies not only that a lesion in one system will produce a bias in the opposite system, but that the bias may be negated either by a repletion of function in the first system (facilitation), or by depletion of function by lesioning, again, or depressing the second system. Utilizing the known fact that posterior singulate lesions produce losses on a CAR (McCleary, 1961; Thomas & Slotnick, 1962), it is interesting that drive deprivation, which produced EEG arousal in the rat (Hockman, 1964), negates the effects of such posterior cingulate lesions (Thomas & Slotnick, 1963). Further, anterior cingulate and septal lesions, which alone produce passive avoidance disruption, are negated in this effect by posterior cingulate lesions (Lubar, 1964; McCleary, 1961). Although medial thalamic nuclei lesions disrupt active avoidance responding (Thompson, 1963; Vanderwolf, 1963), Vanderwolf (1964) has found that additional septal lesions lessen this effect. Consistently, he found septal lesions alone to lessen the acquisition rate on the CAR even though these lesions, when superimposed on medial thalamic lesions, improved performance on the same task. It will be seen after the following section that many drugs could be substituted for the drive effects obtained by Thomas and Slotnick (1963) or the lesion effects of Vanderwolf (1964) or Lubar (1964).

Drug Effects

The production of neural biases by drug action superimposes another dimension of action on the electrophysiological and neurological data described to this point. In many cases, such investigations promise to clarify the activities of systems where these systems anatomically or electrically overlap (cf. Fisher, 1964). In other cases, investigations involving drug effects provide excellent convergence on the brain system biases discussed previously in relation to lesion and recording data. Grossman (1964), for example, found that crystalline implants of cholinergic, but not adrenergic, stimulants will interfere with performance on a CAR when these implants are placed in the medial septal region of the brain. As stated previously, it is quite interesting that electroencephalographic data indicate that highly synchronous hippocampal slow waves occur as a result of cholinergic stimulation (Sailer & Stumpf, 1957), and are blocked by septal lesions (Mayer & Stumpf, 1958), which in turn facilitate performance on a CAR (King, 1958). Using peripheral, rather than central, injections, Chalmers and Erickson (1964) noted that this net cholinergic stimulation also blocks performance on a CAR. Atropine, which blocks cholinergic stimulation (Green, Maxwell, Schindler, & Stumpf, 1960), again released CAR responding. These effects are substantiated in the similar and earlier findings of Pfeiffer and Jenney (1957). Meyers, Roberts, Riciputi, and Domino (1964) have additionally found that cholinergic blocking drugs (atropine and scopalamine) disrupted only the acquisition, but not the retention, of a CAR. These effects were again substantiated in the findings of Herz (1959) and of Bignami (1964). Hearst (1959) also reported the cholinergic blocker scopalamine to disrupt differential responding, to produce an inability to withold responding between trials, and to produce deficits on response extinction. Buresova, Bures, Bohdanecky, and Weiss (1964) have found atropine to affect passive avoidance and extinction only when these reactions were not overlearned. Chlorpromazine, which apparently acts as an adrenergic blocker (Brodie, 1959), has been found to facilitate, or at least not to affect, performance of a passive avoidance (Naess & Rasmussen, 1958; Ray, 1964), and to block performance on a CAR (Chalmers & Erickson, 1964). Bindra and Mendelson (1962) have found, consistent with the notion of bias toward a control system, that chlorpromazine had greater retarding effects on performance where greater overtraining had previously occurred on the task. Amphetamine, which mimics to some degree the adrenergic stimulants (Brodie, 1959), has been found to differentially interfere with the inhibitory component of a discrimination (Cardo, 1959) and to facilitate performance on a CAR (Carlton & Didamo, 1961).

These results tend to indicate that the brain systems underlying control and release functions are biased by cholinergic stimulation and adrenergic blockade (Bindra & Mendelson, 1962; Chalmers & Erickson, 1964; Grossman, 1964; Naess & Rasmussen, 1958; Ray, 1964), and by adrenergic stimulation and cholinergic blockade (Buresova, Bures, Bohdanecky, & Weiss, 1964; Cardo, 1959; Carlton & Didamo, 1961; Hearst, 1959; Herz, 1959; Meyers, Roberts, Riciputi, & Domino, 1964; Pfeiffer & Jenney, 1957), respectively. Such mutual cholinergic-adrenergic inhibition systems have also been reported by Grossman (1964) in his many studies on chemical stimulation of eating and drinking in rats. Since Grossman reported interference with a CAR performance in the same implant (cholinergic) that stimulates drinking, such findings on the production of consummatory responses might provide interesting converging evidence on the anatomical components of the brain systems underlying the control and release functions described in this paper. It will be remembered that neurological evidence (Gerbrandt, 1964) indicated that a circuit comparable to that of Papez (1937) and a frontal to septal to tegmental (Le Gros Clark & Meyer, 1950) pathway each seemed to function in increasing the likelihood of a response in competition with more likely responses. It seems now that this system may be biased by cholinergic stimulation, or by adrenergic blockade. Correspondingly, Fisher and Coury (1962) noted that drinking behavior is mediated by these same characteristically cholinergic Papez and medial forebrain circuits.

While such remarkable support has come from independent investigations on the production of drinking, a systematic attack is needed to elucidate these relations in the same animals. Possible relations between the adrenergic systems of eating and the proposed system for behavioral release are in especial need of such clarification. Unlike one of the interpretations of Grossman's (1964) data, cholinergic stimulation is proposed in the present paper to lead to CAR performance decrements not because of competition with motivated (drinking) behavior. It is hypothesized instead that at sites eliciting eating by adrenergic stimulation, no such competition should result from eating (or drinking). Also, Hernandez-Peon, Chavez-Ibarra, Morgane, and Iaria (1963) did not report such consummatory effects in the cat, but their findings did support the notion of a cholinergic arousal inhibition system running from the septal area to the tegmentum. Again, cholinergic stimulation should lead to blockage of the CAR performance in spite of the lack of competing consummatory responses. Finally, those areas where anatomical overlap of adrenergic and cholinergic systems is known to occur (Grossman, 1962) should be investigated for overlap of the proposed control and release functions thought to be related to these systems. Such investigations should reveal components of the systems of concern in this paper which have heretofore been masked by dimensions of technique.

CONCLUSIONS

In this paper, a descriptive model noting that discrete brain systems determine the release and control of stabilized responses was formed from suggestions and behavioral parameters in the existing literature. Many of the types of "contradiction" between studies in neurology (Moore, 1964; Vanderwolf, 1964), electrophysiology (Adey et al., 1960; Grastyan et al., 1959), and drug research (Buresova, Bures, Bohdanecky, & Weiss, 1964; Chalmers & Erickson, 1964) were resolved by this model. Since the model was formed from existing suggestions in the literature, it might only be expected to differ from the other models by its added generality in the ordering of release and control functions of behavior. In this sense the existing models differ by limiting that which is released or controlled. Response inhibition concepts (McCleary, 1961; Schwartzbaum, Kellicutt, Spieth, & Thompson, 1964) fail to account for the differences which may be obtained on acquisition and retention (Adey et al., 1964; Kenyon, 1962; Meyers et al., 1964). As explanatory constructs, releases and controls over freezing (Thomas & Slotnick, 1963; Vanderwolf, 1964) or motility (Maher & McIntyre, 1960) do not offer clear predictions on acquisition and retention effects. These constructs are also faced with cases where deficits are obtained without changes in freezing or motility (Gerbrandt, 1965; Krieckhaus, 1964), and are relatively species-specific to rat and cat behavioral changes. Release and control of orienting (MacAdam, 1962; Chalmers & Erickson, 1964) or attentive shifting (Gerbrandt, 1964) mechanisms may potentially have the same predictive value as the model set forth in this paper; however, they are at present too easily used to settle interpretations in an ad hoc fashion and differ by laying the causal basis for effects on other correlated effects (epiphenomena?). Again, in the present review the model is intended as a valid description of the events of release and control of behaviors which may be ranked in stability. Perhaps the causal natures of these data will become clarified once such a descriptive model, although it may not be the present one, is agreed upon as fitting all of the relevant facts. (For more extensive theoretical treatments of these data, see the paper by Sachs, 1965, and the book by Fair, 1963.)

REFERENCES

Adey, W. R., Dunlop, C. W., & Hendrix, C. E. Hippocampal slow waves. *Archives of Neurology*, 1960, **3**, 74–90.

Adey, W. R., Sunderland, S., & Dunlop, C. W. The entorhinal area: Electrophysiological studies of its interrelations with the rhinencephalon and the brain stem. *Electroencephalography and Clinical Neurophysiology*, 1957, **9**, 309–324.

Bignami, G. Effects of benactyzine and adiphene on instrumental conditioning in a shuttlebox. *Psychopharmacologia*, 1964, **5**, 264–279.

Bindra, D., & Mendelson, J. Interaction of habit strength and drug effects. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 217–219.

Bitterman, M. E. Techniques for the study of learning in animals: Analysis and classification. *Psychological Bulletin*, 1962, **59**, 81–93.

Bremner, F. J. Hippocampal activity during avoidance behavior in the rat. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 16–22.

Brodie, B. B. Interaction of drugs with norepinephrine in the brain. In O. Krayer (Ed.), *Symposium on catecholamines*. Baltimore: Williams & Wilkins, 1959. Pp. 548–564.

Brutkowski, S. The effect of prefrontal lobectomies on salivary conditioned reflexes in dogs. *Acta Biologica*, 1957, **17**, 327–337.

Brutkowski, S., & Mempel, E. Disinhibition of inhibitory conditioned response following selective brain lesions in dogs. *Science*, 1961, **17**, 327–337.

Buresova, O., Bures, J., Bohdanecky, Z., & Weiss, T. Effect of atropine on learning, extinction, retention, and retrieval in rats. *Psychopharmacologia*, 1964, **5**, 255–263.

Buresova, O., Bures, J., Fifkova, E., Vinogradova, O., & Weiss, T. Functional significance in corticohippocampal connections. *Experimental Neurology*, 1962, **6**, 161–172.

Cardo, B. Action de l'amphetamine dextrogyre et de l'eserine sur un conditionnement de fuite et sur les phenomenes de discrimination. *Journal de Physiologie*, 1959, **51**, 845–860.

Carlton, P. L., & Didamo, Paulene. Augmentation of the behavioral effects of amphetamine by atropine. *Journal of Pharmacology and Experimental Therapeutics*, 1961, **134**, 91–96.

Chalmers, R. K., & Erickson, C. R. Central cholinergic blockade of the conditioned avoidance response in rats. *Psychopharmacologia*, 1964, **6**, 31–41.

Dahl, D., Ingram, W. R., & Knott, J. R. Diencephalic lesions and avoidance learning in cats. *Archives of Neurology*, 1962, 314–319.

Ellen, P., Wilson, A. S., & Powell, E. W. Septal inhibition and timing behavior in the rat. *Experimental Neurology*, 1964, **10**, 120–132.

Elul, R. Regional differences in the hippocampus of the cat: II. Projections of the dorsal and ventral hippocampus. *Electroencephalography and Clinical Neurophysiology*, 1964, **16**, 489–502.

Fair, C. M. *Physical foundations of the psyche.* Middletown, Mass.: Wesleyan Univer. Press, 1963.

Fisher, A. E. Chemical stimulation of the brain. *Scientific American*, 1964, **210**, 60–68.

Fisher, A. E., & Coury, V. N. Cholinergic tracing of a central neural circuit underlying the thirst drive. *Science*, 1962, **138**, 691–693.

Gerbrandt, L. K. Generalizations from the passive and active avoidance distinction. *Psychological Reports*, 1964, **15**, 11–12.

Gerbrandt, L. K. The effects of anteromedial and dorsomedial thalamic lesions on passive avoidance and activity. *Psychonomic Science*, 1965, **2**, 39–40.

Grastyan, E., Lissak, K., Madarasz, I., & Donhoffer, H. Hippocampal electrical activity during the development of conditioned reflexes. *Electroencephalography and Clinical Neurophysiology*, 1954, **17**, 533–557.

Green, J. D., & Arduini, A. Hippocampal electrical activity in arousal. *Journal of Neurophysiology*, 1954, **17**, 533–547.

Green, J. D., Maxwell, D. S., Schindler, W. J., & Stumpf, C. Rabbit EEG "theta" rhythm: Its anatomical source and relation to activity in single neurons. *Journal of Neurophysiology*, 1960, **23**, 403–420.

Grossman, S. P. Effects of adrenergic and cholinergic blocking agents on hypothalamic mechanisms. *American Journal of Physiology*, 1962, **202**, 1230–1236.

Grossman, S. P. Effect of chemical stimulation of the septal area on motivation. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 194–200.

Hearst, E. Effects of scopolamine on discriminated responding in the rat. *Journal of Pharmacology and Experimental Therapeutics*, 1959, **126**, 344–358.

Hernandez-Peon, R., Chavez-Ibarra, G., Morgane, P. J., & Iaria, T. Limbic cholinergic pathways involved in sleep and emotional behavior. *Experimental Neurology*, 1963, **8**, 93–111.

Herz, A. Uber die wirkung von scopolamin, benactyzin und atropin auf das reaktiv verhalten der ratte. *Archiv für Experimentelle Pathologie und Pharmakologie*, 1959, **236**, 110–112.

Hockman, C. H. EEG and behavioral effects of food deprivation in the albino rat. *Electroencephalography and Clinical Neurophysiology*, 1964, **17**, 420–427.

Hunt, H. F., Diamond, I. T., Moore, R. Y., & Harvey, J. A. Some effects of hippocampal lesions on conditioned avoidance behavior in the cat. *Proceedings of the XVth International Congress of Psychology*, 1957, **8**, 203–204.

Isaacson, R. L., Douglas, R. L., Douglas, R. J., & Moore, R. Y. The effect of radical hippocampal ablation on acquisition of avoidance response. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 625–628.

Isaacson, R. J., & Wickelgren, W. O. Hippocampal ablation and passive avoidance. *Science*, 1962, **138**, 1104–1106.

John, E. R., & Killam, K. F. Electrophysiological correlates of avoidance conditioning in the cat. *Journal of Pharmacology and Experimental Therapeutics*, 1959, **125**, 252–274.

Kaada, B. R., Rasmussen, E. W., & Kveim, O. Impaired acquisition of passive avoidance behavior by subcallosal, septal, hypothalamic, and insular lesions in rats. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 661–670.

Kenyon, J. The effect of septal lesions upon motivated behavior in the rat. Unpublished doctoral dissertation, McGill University, 1962.

Kimble, D. P. The effects of bilateral hippocampal lesions in rats. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 273–283.

Kimura, D. Effects of selective hippocampal damage on avoidance behavior in the rat. *Canadian Journal of Psychology*, 1958, **12**, 213–238.

King, F. A. Effects of septal and amygdaloid lesions on emotional behavior and conditioned avoidance responses in the rat. *Journal of Nervous and Mental Disease*, 1958, **126**, 57–63.

Krieckhaus, E. E. Decrements in avoidance behavior following mammillothalamic tractotomy in cats. *Journal of Neurophysiology*, 1964, **27**, 753–767.

Krieckhaus, E. E., Simmons, H. J., Thomas, G. J., & Kenyon, J. Septal lesions enhance shock avoidance behavior in the rat. *Experimental Neurology*, 1964, **9**, 107–113.

Le Gros Clark, W. E., & Meyer, M. Anatomical relationships between the cerebral cortex and the hypothalamus. *British Medical Bulletin*, 1950, **6**, 341–345.

Locke, S., Angevine, J. B., & Yakovlev, P. I. Limbic nuclei of the thalamus and connections of limbic cortex: VI. Thalamocortical projections of lateral dorsal nucleus in cat and monkey. *Archives of Neurology*, 1964, **11**, 1–12.

Lubar, J. F. Effect of medial cortical lesions on the avoidance behavior of the cat. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 38–46.

MacAdam, D. W. Electroencephalographic changes and classical aversive conditioning in the cat. *Experimental Neurology*, 1962, **6**, 357–371.

McCleary, R. A. Response specificity in the behavioral effects of limbic system lesions in the cat. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 605–613.

McCleary, R. A., Jones, Caroline, & Ursin, H. Avoidance and retention deficits in septal cats. *Psychonomic Science*, 1965, **2**, 85–86.

Maher, B. A., & McIntire, R. W. The extinction of the CER following frontal ablation. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 549–552.

Mahut, Helen, & Cordeau, J. P. Spatial reversal deficit in monkey with amygdalohippocampal ablations. *Experimental Neurology*, 1963, **7**, 426–434.

Mayer, C., & Stumpf, C. Die physostigminwirkung auf die hippocampus-tätigkeit nach septum läsionen. *Archiv für Experimentelle Pathologie und Pharmakologie*, 1958, **234**, 490–500.

Meyers, B., Roberts, K. H., Riciputi, R. H., & Domino, E. F. Some effects of muscarinic cholinergic blocking drugs on behavior and the electrocorticogram. *Psychopharmacologia*, 1964, **5**, 289–300.

Moore, R. Y. Effects of some rhinencephalic lesions on retention of conditioned avoidance behavior in cats. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 65–71.

Morrell, F. An anatomical and physiological analysis of electrocortical conditioning. *Proceedings of the 1st International Congress of Neurological Science*, Brussels, 1957, 377–391.

Naess, K., & Rasmussen, E. W. Approach-withdrawal responses and other specific behavior reactions as screening tests for tranquilizers. *Acta Pharmacologica et Toxicologica*, 1958, **15**, 99–114.

Nauta, W. J. H., & Whitlock, D. G. An anatomical analysis of the nonspecific thalamic projection system. In J. F. Delafresnaye (Ed.), *Brain mechanisms and consciousness*. Springfield, Ill.: Charles C Thomas, 1954. Pp. 81–116.

Papez, J. W. A proposed mechanism of emotion. *Archivio di Psicologia, Neurologia e Psichiatria*, 1937, **38**, 725–745.

Pavlov, I. P. *Lectures on conditioned reflexes*. New York: International Publishers, 1928.

Petsche, H., & Stumpf, C. Topographic and topоscopic investigations on origin and spread of the regular diencephalo-hippocampal arousal activity in the rabbit. *Electroencephalography and Clinical Neurophysiology*, 1960, **12**, 589–600.

Pfeiffer, C. C., & Jenney, Elizabeth H. The inhibition of the conditioned response and the counteraction of schizophrenia by muscarinic stimulation of the brain. *Annals of the New York Academy of Science*, 1957, **66**, 753–764.

Ploog, D. W., & MacLean, P. D. On functions of the mammillary bodies in the squirrel monkey. *Experimental Neurology*, 1963, **7**, 76–85.

Powell, E. W. Septal efferents revealed by axonal degeneration in the rat. *Experimental Neurology*, 1963, **8**, 406–422.

Pribam, K. H., & Kruger, L. Functions of the "olfactory brain." *Annals of the New York Academy of Science*, 1954, **58**, 109–138.

Ray, O. S. The effect of central nervous system depressants on discrete trial approach-avoidance behavior. *Psychopharmacologia*, 1964, **6**, 96–111.

Rosvold, H. E., & Szwarcbart, M. K. Neural structures involved in delayed-response performance. In J. M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior*. New York: McGraw-Hill, 1964. Pp. 1–15.

Sachs, E. Dissociation of learning in rats and its possible similarities to dissociative states in man. *Annual Journal of Psychology*, in press.

Sailer, S., & Stumpf, C. Beeinflussbarkeit der rhinencephalen tätigkeit des kaninchens. *Archiv für Experimentelle Pathologie und Pharmakologie*, 1957, **231**, 63–77.

Schwartzbaum, J. S., Kellicutt, M. H., Spieth, T. M., & Thompson, J. B. Effects of septal lesions in rats on response inhibition associated with food-reinforced behavior. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 217–224.

Stanley, W. C., & Jaynes, J. The function of the frontal cortex. *Psychological Reviews*, 1949, **56**, 18–32.

Teitelbaum, H. A comparison of effects of orbitofrontal and hippocampal lesions upon discrimination learning and reversal in the cat. *Experimental Neurology,* 1964, **9**, 452–462.

Teitelbaum, H., & Milner, P. Activity changes following partial hippocampal lesions in rats. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 284–289.

Thomas, G. J., Fry, W. J., Fry, F. J., Slotnick, B. M., & Krieckhaus, E. E. Behavioral effects of mammilothalamic tractotomy in cats. *Journal of Neurophysiology,* 1963, **26**, 857–876.

Thomas, G. J., & Slotnick, B. Effects of lesions in the cingulum on maze learning and avoidance conditioning in the rat. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 1085–1091.

Thomas, G. J., & Slotnick, B. Impairment of avoidance responding by lesions in cingulate cortex in rats depends on food drive. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 959–964.

Thompson, R. Thalamic structures critical for retention of an avoidance conditioned response in rats. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 261–267.

Thompson, R., & Langer, Susan K. Deficits in position reversal learning following lesions of the limbic system. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 987–995.

Vanderwolf, C. H. The effect of medial thalamic lesions on previously established fear-motivated behavior. *Canadian Journal of Psychology,* 1963, **17**, 183–187.

Vanderwolf, C. H. Effect of combined medial thalamic and septal lesions on active-avoidance behavior. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 31–37.

Vanderwolf, C. H., & Heron, W. Electroencephalographic waves with voluntary movement. *Archives of Neurology,* 1964, **11**, 379–384.

Webster, D. B., & Voneida, T. J. Learning deficits following hippocampal lesions in split-brain cats. *Experimental Neurology,* 1964, **10**, 170–182.

Winnick, Wilma A., & Hunt, J. McV. The effect of an extra stimulus upon strength of response during acquisition and extinction. *Journal of Experimental Psychology,* 1951, **41**, 205–215.

(Received December 28, 1964)

Psychological Bulletin
1965, Vol. 64, No. 2, 124–150

SELECTIVE ATTENTION IN ANIMAL DISCRIMINATION LEARNING[1]

N. J. MACKINTOSH
Oxford University

Two classes of experiment on the role of attention in discrimination learning are reviewed: (a) Investigations of the effect of attention on the amount learned about different cues have been interpreted as disproving noncontinuity theory (according to which animals attend to only one cue at a time). The fact that animals learn something about a second cue, however, does not prove that attention has no effect on learning, and more recent evidence shows that it does. (b) If animals do not automatically attend to all cues, part of what they must learn in order to solve a discrimination problem is to attend to the relevant cue. Experiments on the acquired distinctiveness of cues, transfer along a continuum, and reversal learning provide evidence for the importance of such classificatory learning.

The concept of attention has occupied a rather humble place in the history of psychology. Seventy-five years ago William James (1890) wondered how it could be that "so patent a fact as the perpetual presence of selective attention has received hardly any notice from psychologists of the English empiricist school [p. 402]." Later, it is true, among the structuralists, the related notion of "attensity" played a more important role; but with the rise of behaviorism, attention was once more banished from psychological discourse, and even the Gestalt psychologists (less inclined than the early behaviorists to prejudge the value of a concept from its introspective overtones) had little room in their system for a selective determinant of perception. One of the more valuable features of the impact of information theory on psychological thinking has been in its insistence on the importance of a number of input variables in the determination of behavior, and the concept of limited channel capacity, as has been pointed out by Broadbent (1958), inevitably leads to a consideration of the importance of selective or filtering processes in perception.

Animals (particularly lower animals) have nervous systems of limited size and therefore of limited capacity for processing and storing information. Thus they are confronted with the problem of selection. At some stage they must discard irrelevant or redundant information so as not to interfere with the storage of important information. This line of argument would seem to provide a general rationale for postulating, as Broadbent does, the existence of filtering devices in the nervous system; and if this approach is justified, it is undoubtedly of the first importance for general behavior theory. To put it at its simplest, if animals do not respond to all features of their stimulus input, then a sharp distinction must be drawn between the physical stimuli impinging on an animal in any given situation and the effective stimulus which controls the animal's behavior in that situation. Failure to consider this distinction might lead to explanations of behavior being offered that are at best incomplete and at worst totally misconceived. As Prokasy and Hall (1963) have recently pointed out, this failure may well invalidate much of the theoretical approach to the problem of stimulus generalization.

It may be granted, then, both that there are plausible grounds for postulating a central mechanism of attention, and that the question of attention is an important one; it remains to consider the nature of the experimental evidence bearing on these issues. This review makes no attempt at an exhaustive coverage of the field; it is confined to experiments which have investigated the role of attention in discrimination learning and, even within this area, is largely concerned with learning in subhuman animals (although some work with humans will be mentioned). Animal discrimination learning is one of the few topics where the concept of attention has been applied in the past. Lashley (1929) was perhaps the first investigator to raise the possibility that animals solving discrimination problems might not attend to all stimuli impinging on their receptors, and this suggestion subsequently provided one of the points at issue in the controversy between continuity and noncontinuity theories of discrimination learning. This controversy contained several elements over and above the question of the role of attention in discrimination learning. Among them were: (*a*) hypothesis behavior during discrimination learning—this has recently been the subject of several mathematical formulations (Levine, 1959; Restle, 1962); (*b*) the rate at which learning occurs—as the name implies, noncontinuity theories usually regard learning as occurring very rapidly, perhaps in a single trial (Estes, 1960, 1964); and (*c*) the nature of the stimulus in discrimination learning—whether animals respond to the relationship between the positive and negative stimuli or to their absolute values.

---

[1] This paper was written as part of a project on Stimulus Analyzing Mechanisms, financed by the American Office of Naval Research (Contract N62558-3937) and the Nuffield Foundation. I am grateful to a number of investigators who have allowed me to see prepublication copies of papers; to J. Mackintosh, E. M. Macphail, and N. S. Sutherland for their comments on an earlier draft, and finally to E. P. Lovejoy and, in particular, to N. S. Sutherland who are responsible for many of the ideas in this review.

Important as these three issues are in their own right, they are not of present concern. It is arguable that the first two at any rate derive from what is here regarded as the central contention of noncontinuity theory—that animals respond selectively—but they are certainly not necessary consequences of this view. It is difficult to see any logical connection between the question of the role of attention in discrimination learning and the question of the absolute or relational nature of the stimulus controlling behavior. The present review, therefore, is concerned exclusively with the central question of the role of attention in discrimination learning. Two broad classes of experiment are discussed: first, those designed to investigate the effect of attention on the amount learned about any given cue; secondly, those concerned with the possibility that in order to solve a discrimination problem, animals must learn both to attend to the relevant cue and to establish appropriate choice responses. Finally, a number of two-stage theories of discrimination learning are briefly discussed, and an attempt is made to bring out some of their similarities and differences.

EXPERIMENTS ON THE SELECTIVITY OF LEARNING

The most obvious implication of the concept of attention is that animals should not learn equally about all the stimuli impinging on their receptors. Classically, three types of experiment have been generated by this point of view: (*a*) those on the presolution period, (*b*) those on what this reviewer will call incidental cue learning, and (*c*) those on single-stimulus pretraining. The plan will be to review earlier results, in their light to reject one version of noncontinuity theory as originally formulated, to outline the elements of an alternative noncontinuity position, and, finally, to show how this reformulated theory is compatible both with early results and with more recent experiments.

The Presolution Period

There is no reason for dealing at length with this set of experiments, since it was unnecessary to perform them in order to disprove an extreme noncontinuity position. The rationale for the experiments rested on the observed fact that rats, trained by a noncorrection procedure on a simultaneous visual discrimination, tend, while performing at a chance level of accuracy, to select a particular position consistently; in other words, they "adopt a spatial hypothesis" or "display a position habit." The breaking of this position habit usually precedes perfect performance by a relatively small number of trials. From this observation (which was never in dispute), Krechevsky (1938) inferred that, while responding to one stimulus, the rat "learns (wrongly perhaps) something about the significance of this particular stimulus . . . but does not learn anything about the 'correctness' or 'wrongness' of the to-be-finally-learned set of discriminanda [p. 111]." It is not certain that this statement was ever a fair representation of Krechevsky's views; in the same article there are other statements putting forward a much more moderate position. However, the extreme position implied by this statement was in fact accepted as an integral part of noncontinuity theory, and in the remainder of this section it will occasionally, for brevity's sake, be referred to as Krechevsky's position. The particular statement quoted here seems to rest on two unrelated assumptions: (*a*) that consistent selection of a particular position indicates that the subject is attending to position cues, and (*b*) that attention to position cues precludes the possibility of learning anything about other cues. The first assumption can perhaps be safely accepted. But there is no a priori justification for the second assumption, and evidence against it is provided by two simple observations of the behavior of rats learning simultaneous visual discriminations.

1. Mahut (1954) reported that while rats were still responding to position (i.e., during the presolution period) the latencies of their responses to the positive and negative stimuli (S+ and S−) began to draw apart. Although they selected S+ no more frequently than S−, they responded to it significantly faster. The writer has also observed this effect.[2] Thus the use of a (more sensitive) latency measure reveals that a preference may be established for S+ over S−, when a choice measure indicates total absence of learning.

2. If, while the subjects are consistently responding to position, they learn nothing about the relevant cue, it follows that their first response away from their preferred position should on the average be as often to S− as to S+. In fact, of 40 subjects trained by the writer on a black-white door discrimination in a jumping stand, 30 developed consistent position habits (selected one position for at least 20 consecutive trials), and of these 30, only one first broke its position habit by selecting S− on its nonpreferred side.

It is clear that rats will both respond systematically to position and learn something about the relevant cue at one and the same time. It follows that Krechevsky's second assumption is false and that the series of experiments designed to test this extreme noncontinuity position was redundant.

In fact, of course, the majority of these experiments have yielded results incompatible with Krechevsky's point of view. The design of these experiments was to compare two groups in their rate of learning a discrimination A'-A''.[3] One group was trained throughout with A' positive, the other (reversal) group was initially trained with A'' positive, but was reversed to A' positive while still responding at chance (before the end of the presolution period). Results contradicting Krechevsky's noncontinuity theory have been obtained in the following experiments of this design with rats: Ehrenfreund (1948, Experiment II), Gatling (1951), Krechevsky (1938, Group III), McCulloch and Pratt (1934), Ritchie, Ebeling, and Roth (1950), Spence (1945), and in one experiment with human subjects, Walk (1952). Four experiments have shown no sign of relevant cue learning during the presolution period: Bower and Trabasso (1963), Bradley (1961), Ehrenfreund (1948, Experiment I), and Krechevsky (1938, Group II). Bradley used human subjects but, instead of a reversal group, had a group for whom the relevant cue was not introduced until after 10 trials. The fact that these subjects learned the problem as fast as a control group need not show that the control group failed to attend to the revelant cue during the first 10 trials, but only that the attention of the experimental group was attracted to the relevant cue by its novelty. In Ehrenfreund's experiment with rats, the stimuli were placed on the top half of the doors of a jumping stand, and it can reasonably be claimed that the absence of presolution-period learning was due to the animals having failed to *look at* the relevant stimuli. The results, therefore, can hardly be adduced in support of noncontinuity theory.

---

[2] Unpublished observations.

[3] Here, and elsewhere in this article, the following notations are used: A, B, etc., are used to refer to cues, or stimulus dimensions such as brightness, size, etc.; A', A'' are used to refer to particular stimuli embodying particular values of these dimensions (for example, A' black, A'' white; if A' signifies a black square and A'' a white square, then $A_{,}$, might signify a black diamond, that is, a stimulus differing from A' along a different dimension). In the phase, "a discrimination between A' and A''," the first stimulus is positive, the second negative.

There remain Krechevsky's results with rats, and those of Bower and Trabasso with college students. Krechevsky used a very difficult discrimination, Bower and Trabasso either a complex one or a very small number of reversed pretraining trials. Although, therefore, as both Bower and Trabasso (1963) and Broadbent (1961) have suggested, it is possible that selective attention may only operate in complex situations, it is also possible that these experiments were biased in favor of producing negative results, since on any theory the amount of learning to be expected in these situations would be extremely small. This, of course, raises the crucial difficulty with this experimental design—the problem of attempting to prove the null hypothesis. There have been seven convincing demonstrations that reversal of stimuli during the presolution period can significantly retard subsequent discrimination learning; in two papers only has evidence failing to support this generalization been found. This evidence is hardly sufficient basis for any theory.

These experiments, therefore, could never have proved noncontinuity theory to be correct. Noncontinuity theorists set themselves the task of demonstrating absence of learning. It was always open to an opponent to claim, of any particular experiment which found no difference between reversed and nonreversed subjects, that insufficient numbers had been run or that an insensitive measure of learning had been used. Since, as will be pointed out later, the experiments do not even provide good evidence that continuity theory is correct, little more needs to be said about them.

Incidental Cue Learning

The logic of the presolution-period experiments was derived from the way in which rats solve simultaneous discrimination problems. Since rats tend to respond systematically to position in the early stages of training, it was not entirely implausible to suggest that they are attending predominantly to position, and less to the relevant cue. Lashley (1942) pointed out that on noncontinuity theory it should also be possible to preset an animal's attention by specific experimental procedures. An animal that has been trained on a size discrimination between a large and small circle might be assumed to be attending to the cue of size. If now a large triangle is substituted for the large circle and training is continued, an incidental cue (shape) is being added to the primary cue (size). But if the subject attends exclusively to the primary cue, it should learn nothing about the secondary cue. (It is worth noting that attention here is not defined in terms of systematic responses, since the subject in this situation responds just as systematically to shape as to size.)

Although fewer experiments have been performed on incidental cue learning than on the presolution period, their results are hardly more favorable to noncontinuity theory as originally formulated. It is true that Lashley, when he performed an experiment such as the one outlined above, claimed to find no evidence that incidental learning had occurred, but he used one insensitive and one quite inappropriate measure of learning. In his first test of whether the subjects had learned about the shape of the stimuli, he gave transfer tests with a triangle and a circle of equal size and found random responding. But, as noted by Sutherland (1959), transfer tests with rats are too insensitive to reveal small amounts of learning, since (as happened on this occasion) subjects revert to position habits. His second test involved opposing the primary and incidental cues; the fact that the primary cue controlled behavior is neither surprising, nor is it good evidence for the total ineffectiveness of the secondary cue.

Blum and Blum (1949) performed an experiment similar to Lashley's and found some incidental learning to occur, but although they eliminated the undesirable features of Lashley's design, they incorporated further inadequacies of their own. Rats were trained on the large-small circle discrimination; once they had learned, a large erect triangle was substituted for the large circle, and a small inverted triangle for the small circle, and training was continued. When given tests with triangles of equal size, the subjects continued to select the erect triangle. Now it is generally accepted that in the jumping stand the rat's behavior is controlled more by the bottom half of the stimuli than by the top half (Lashley, 1938). There is surely, therefore, every reason to believe that of two equally sized triangles, the erect (with a long base line) will look larger to the rat than the inverted (with a point as base).

Fortunately, there have been two more recent and entirely satisfactory experiments—those of Bitterman and Coate (1950) and Hughes and North (1959). Each of these shows beyond any doubt that animals will learn something about an incidental cue (even if it is only partially, not fully, correlated with reward). The conclusion again seems certain: In Lashley's and Krechevsky's formulations, noncontinuity theory has been disproved.

Single Stimulus Pretraining

The third experimental situation was originally devised by Lashley and Wade (1946). If the subject is trained to respond to a single stimulus *A′*, and subsequently learns the discriminations *A′*-*A″*, continuity theory will normally predict faster learning of the test discrimination if *A′* is positive (nonreversal) than if *A′* is negative (reversal). Lashley and Wade denied this. Their argument rested (as did Krechevsky's on the presolution period) on two unrelated assumptions: (*a*) that during training with *A′* alone, a sloping gradient of generalization along the *A′*-*A″* dimension would only be formed if the subject classified *A′* as falling at some point along the dimension—that is, attended to this feature of *A′*; and (*b*) that the subject would not attend to a particular dimension unless given discrimination training between two points along the dimension. Hence no generalization gradient would be formed after single-stimulus pretraining. It is important to see that these two assumptions are independent. A demonstration of sloping gradients of generalization following single-stimulus pretraining might disprove the second assumption but leave the first intact.

Lashley and Wade briefly reported a number of experiments using variously rats, monkeys, and chimpanzees, in none of which were significant differences between reversed and nonreversed groups found (note again the use of an experimental design which can only show noncontinuity theory to be correct by proving the null hypothesis). Two of their rat experiments were repeated by Grice (1948, 1951), who in each case found significant differences in favor of the nonreversed groups. Grice's results are frequently quoted (Kimble, 1961, pp. 369–371; Osgood, 1953, p. 449) as providing decisive disproof of Lashley and Wade's position, so it is worth showing that his experimental design is invalid. For the sake of illustration we need only consider his first experiment, although the same conclusions apply to the second. Mackintosh (1965d) trained two groups of rats to approach a white circle, and avoid a black door; for one group (nonreversal) the circle was 8 cm. in diameter, for the second (reversal) group a 5-cm. circle was used. Both groups then learned to approach the 8-cm. circle and avoid the 5-cm. circle, and the nonreversal group learned faster than the reversal group. But when a third and fourth group were trained, the third initially to approach the 5-cm. circle, the fourth with the 8-cm. circle, and then were both trained to approach the 5-cm. circle and avoid the 8-cm. circle, both groups took nearly three times as long to learn this discrimination as did the equivalents of Grice's two groups. Yet according to Grice's analysis, the third group was tested on a nonreversal. These results become intelligible as soon as one sees that in original training all four groups were learning to approach the brighter of two stimuli (a black door with a white circle) and to avoid the darker stimulus (a black door). Grice's two groups were then trained to approach the larger (brighter) of two white circles and therefore learned a nonreversal, while the third and fourth groups were trained to approach the smaller (darker) of two circles and therefore were reversed. Thus Grice's (and Lashley and Wade's) experimental situation involves the successive learning of two brightness discriminations, and the results are not a test of the effect of single-stimulus pretraining at all.

There remains, however, a large number of relevant experiments, which for the first time are not unanimously embarrassing to noncontinuity theory. Although any experiment (e.g., those utilizing pigeons and the spectral dimension, following Guttman and Kalish, 1956) which succeeds in demonstrating orthodoxly sloping gradients seems to provide prima facie evidence against Lashley and Wade's position, there have been an appreciable number in which no evidence for a gradient of generalization could be obtained: (*a*) Between them, Jensen and Cotton (1961) and Thompson (1962) tested for size generalization in 413 rats without finding any gradient, (*b*) McCain and Garrett (1964) found no gradient of brightness generalization in 145 rats, (*c*) Jenkins and Harrison (1960) could find no evidence of a gradient of auditory frequency generalization in pigeons. Also relevant here are those experiments which are taken to show that in order for a neutral stimulus to be established as a secondary reinforcer, it must be a discriminative stimulus. Tests for secondary reinforcement may be regarded as tests of generalization, depending as they do either on showing generalization decrement effects in extinction, or preferences in a choice situation; thus, the results of Ferster (1951), McGuigan and Crockett (1958), and Notterman (1951) may all be taken as showing the importance of discrimination training for producing gradients of generalization.

Heinemann and Rudolph (1963) have recently suggested that even in those experiments ostensibly showing gradients in the absence of discrimination training, there may well have been an inadvertent element of discrimination involved. They noted that pigeons show orthodoxly sloping gradients of spectral generalization after being trained to peck at a key of a given wavelength, but fail to show gradients of auditory frequency generalization after being trained to peck at a key in the presence of a continuously sounded tone of given frequency. They suggested that the former situation is more likely to introduce an element of discrimination training than the latter, since in the former the positive stimulus will only be perceived by the subject when the subject is engaged in key pecking, and therefore likely to be rewarded, while in the latter situation the positive stimulus will be perceived by the subject whether or not the subject is key pecking. They argued that the probability of discrimination training occurring when a visual stimulus is used would vary inversely with the size of the stimulus concerned—the larger the stimulus the more likely the subject would be to see it when not pecking. Their results showed that in accord with this suggestion the slope of a generalization gradient after single-stimulus training to a given wavelength also varied inversely with the area of the stimulus.

It may be that this ingenious suggestion cannot account for all instances of generalization gradients; but it is clear that the occurrence of discrimination training does have a marked effect on generalization, and to this extent it seems that Lashley and Wade (1946) were right. But Lashley and Wade insisted not only on the general importance of discrimination training, but also that such training should be between two points along the same dimension as was to be used in generalization testing. It is true that if such relevant discrimination training is given, the change in slope of the gradient is very considerable—and greater than can reasonably be explained in terms of the summation of excitatory and inhibitory tendencies (Jenkins & Harrison, 1962); but in the majority of the studies cited above, discrimination has usually been between the presence and absence of the stimulus. On the other hand, when discrimination training is given along an entirely different dimension, it has been shown to be totally ineffective in producing generalization gradients (MacCaslin, Wodinsky, & Bitterman, 1952).

The evidence from generalization experiments, therefore, is perhaps moderately favorable towards noncontinuity theory. Nevertheless any verdict based on the total evidence so far reviewed can only be unfavorable. There is no possible doubt that some learning about the relevant cue does occur during the presolution period; there is equally no doubt that rats learn something about a second, incidental cue when it is paired with a primary cue. It is, on the whole, unlikely that the very strong claim made by Lashley and Wade (1946) about generalization can really be substantiated. On the other hand, some of the generalization experiments (e.g., MacCaslin et al., 1952) are by no means easy to reconcile with continuity theory, and it looks as if neither theory is in a very satisfactory position. This conclusion may seem unwarranted to those who have regarded the early experiments (especially those on the presolution period) as good evidence in favor of continuity theory. It is the thesis of the next section that of not one of the experiments so far mentioned can this be said.

REFORMULATION OF NONCONTINUITY THEORY AND SOME NEW EXPERIMENTAL EVIDENCE

The early experiments may have disproved Lashley and Krechevsky's noncontinuity theory; but this would only mean that they have thereby proved continuity theory, if the two theories are not only mutually incompatible, but also cumulatively exhaustive. This they clearly are not. A crude representation of the two positions would be that noncontinuity theory states that animals attend to, and therefore learn about only one cue at a time, while continuity theory states that animals learn equally about all cues impinging on their receptors. Stated thus it is obvious that there are alternative positions; for example, animals might simply attend to some cues more than to others, so that if the subject is attending predominantly to Cue *A*, it will learn more about Cue *A* than about Cue *B*—and less about Cue *B* than it would have done had it not been attending to Cue *A*. But there is no need to insist that the subject either learns everything or learns nothing about a given cue.

Although in an empirical sense this may appear to be a compromise position between the two extreme poles of continuity and noncontinuity theories as categorized above, the theoretical position that emerges clearly has more in common with noncontinuity theory than it has with continuity theory: Emphasis is still placed on a central process of attention which determines (although not now in an all-or-none manner) which stimuli out of all those impinging on it will prepotently control the animal's behavior. In fact, as already mentioned, Krechevsky (1938) himself, in the same article as that from which the extreme noncontinuity statement quoted above was taken, specifically envisaged this possibility:

> an animal might be reacting to an "irrelevant" set of discriminanda, as well as to the relevant set, and would thus be learning *something* [italics added] about the final solution even while seeming to be carrying out an "irrelevant" hypothesis. . . . However . . . since, during the so-called presolution period, the animal does spend some time responding to "irrelevant" discriminanda, the residual neurological effect of a "rewarded" or a "punished" response *is not the same* [italics added] during this period as during the period the animal is paying attention to the relevant discriminanda [pp. 108–109].

This position, therefore, which we may call a modified noncontinuity theory, states simply this: that animals do not classify their stimulus input with equal effectiveness in all possible ways at once, and it should therefore be possible to influence what an animal attends to by appropriate training procedures. The experiments so far reviewed did not provide a test of this position. The finding that animals learn something about the relevant cue during the presolution period shows that in this situation they do not classify their input in only one way at a time; but (as pointed out by Sutherland, 1959) it does not answer the question whether animals responding systematically (i.e., attending) to position learn *less* about the relevant cue than other animals. None of the early experiments, in fact, tested the possibility that attention might have *some* effect on learning; they have all been concerned with investgating whether animals either attend to a given cue, or learn nothing at all about it. Thus the results of these experiments do not contradict expectations of modified noncontinuity theory, and by the same token do not provide evidence in support of continuity theory. For according to continuity theory, not only should animals learn *something* about, for example, visual cues when attending to spatial cues, they should learn *as much* about the visual cues as they would have done had they not been attending to the spatial cues. According to continuity theory there is no such thing as a mechanism of attention over and above peripheral orienting responses, no distinction between primary and incidental cues, and no possibility, therefore, of demonstrating a difference in the amount learned about a given cue depending on the direction of the animal's attention.

It is clear that a different type of experimental design is called for, one that will test for differences in the degree of learning of, for example, primary and incidental cues, or discover whether animals classify their input equally effectively in all ways at once. The modified noncontinuity theory implies that the more likely an animal is to classify its input in one way, the less likely it will be to classify it in another. The most direct test of this assumption is to train animals on a discrimination with two relevant cues, for example, a white horizontal rectangle versus a black vertical rectangle, and once they have reached criterion, to use transfer tests to discover how much they have learned about each cue separately. The hypothesis predicts that the more any particular subject has learned about one of the cues (e.g., brightness), the less it will have learned about the other (orientation). This implies that there should be a negative correlation between individual subject's scores on one cue and their score on the other, and this prediction has been confirmed (Sutherland & Mackintosh, 1964). Reynolds (1961) has reported results from two pigeons that point in the same direction.

Animals do not, therefore, learn about all cues equally, and this suggests that it should be possible to manipulate an animal's attention so as to affect how much it will learn about different cues. Consider the following experimental situation: Rats were trained for 60 trials on this same white horizontal/black vertical discrimination, and tested for the amount learned about orientation. Pretrained on a successive brightness discrimination (i.e., to attend to brightness cues), orientation became a secondary, incidental cue, and although animals given this pretraining learned something about orientation, they learned significantly less than animals not so pretrained (Mackintosh, in press). Similar results have been obtained with octopuses by Sutherland, Mackintosh, and Mackintosh (1965), and with human beings by Eckstrand and Wickens (1954). Thus attention to one cue does not prevent an animal learning anything about a second cue, but it does reduce the amount learned.

These experiments are most directly related to the problem originally raised by Lashley (1942) in his experiment on incidental cue learning, but they also seem relevant to studies on the presolution period. In both areas continuity theory in effect argues that attention will not influence the amount learned about a given cue; these experiments show precisely that it is possible to preset an animal's attention, and thereby to influence the amount learned about various cues. It turns out to be possible to demonstrate the equivalent effect in the area of single-stimulus pretraining.

Lashley and Wade's (1946) account of generalization seemed to rest on two assumptions: the necessity of appropriate classification for the subsequent appearance of a gradient of generalization, and the necessity of appropriate discrimination training for the occurrence of such classification. The experiments reviewed in the previous section were entirely concerned with the second of these assumptions, but as was pointed out the two assumptions are in fact independent, and it is of equal interest to test the first. When rats were trained to jump to a single rectangle of given orientation and brightness, they showed gradients of generalization along both orientation and brightness dimensions (Mackintosh, 1965a). They therefore classified a single stimulus along several dimensions at once (see also, Butter, 1963), a result that can only with difficulty be reconciled with Lashley and Wade's position. However, pretraining subjects on a successive brightness discrimination in the above situation not only produced a significantly steeper brightness generalization gradient, but also a significantly *less* steep gradient along the orientation dimension (orientation pretraining had precisely the opposite effect). Thus rats trained to classify along one dimension are less likely to classify along another than rats given no special training.

Further Variables Affecting Attention

These experiments provide direct support for the modified noncontinuity theory suggested earlier, and equally direct evidence against continuity theory. Having established that the amount learned about any cue can be affected by attention, the next step would seem to be to investigate what variables increase or decrease the range of attention. Such an investigation, if successful, would provide further evidence against continuity theory, which must predict that incidental learning will only be influenced by those factors such as number of trials, or delay of reward, which are important in any learning. This point was first recognized by Bruner, Matter, and Papanek (1955). They trained rats under high or low drive, for 30 or 100 trials on a brightness discrimination in a 4-unit maze, then gave 20 trials with a spatial cue also made relevant, and finally tested for the amount learned about this incidental spatial cue. They found that both increases in drive level and the amount of prior training decreased incidental cue learning.

Two further variables have been definitely identified. Mackintosh (in press) trained rats on an orientation discrimination between a gray horizontal and a gray vertical rectangle. After learning, the brightness of one rectangle was changed to black, of the other to white; training was continued with this incidental brightness cue, and the subjects were finally tested for the amount learned about it. Two factors had a significant effect: (*a*) The greater the difficulty of the original orientation discrimination, the more the subjects picked up about brightness; (*b*) If the rectangles were changed in one step from mid-gray to black or white, the subjects learned more about brightness than if the change was made gradually in a number of small steps. The effect of problem difficulty has also been shown by Terrace (1963): Pigeons were trained on a color discrimination; subsequently, a second cue (orientation) was made relevant, and the subjects were finally tested on this cue alone. If the original colors were faded away gradually before the final stage, subjects made no errors on the orientation cue in isolation; but if the color difference was abolished abruptly, a number of errors followed. Gradual fading out of the color cue would correspond to making the color discrimination more difficult.

At least two experiments on single-stimulus pretraining have shown that problem difficulty is an important determinant of the slope of generalization gradients along a different dimension. MacCaslin et al. (1952) and Perkins, Hershberger, and Weyant (1959) have tested for generalization between A' and $A_{''}$ after discrimination training between A' and A''. Increasing the difficulty of the A'-A'' discrimination produced a significantly steeper gradient of generalization between A' and $A_{''}$. Such results imply that the greater the difficulty of the relevant discrimination, the more likely subjects are to attend to other, irrelevant features of the stimuli, and therefore the more likely they are to classify the stimuli along other dimensions (including that subsequently utilized for the generalization test).

The explanation of the results reviewed in this section has not been fully worked out. There has been a temptation to state (Easterbrook, 1959) merely that such and such a factor "narrows the range of cue utilization," or that under certain circumstances rats do not "actively inspect" the stimuli; and such statements, of course, are merely descriptive.

EXPERIMENTS ON TWO-STAGE LEARNING MODELS

The Importance of Classificatory Learning

At this stage there should be little doubt that animals do not classify the stimulus input equally in all ways at once. They react selectively, and this selectivity can be affected by various training procedures. But this seems to have an important implication, for if animals do not automatically classify their input simultaneously in all possible ways, it is surely possible that an animal confronted with a discrimination problem will not necessarily classify the stimuli appropriately; it might register the position of the stimuli, their orientation and their shape, when in fact the problem is a brightness discrimination. From this it follows that part of what a subject must learn in order to solve a discrimination is precisely to attend to the relevant stimulus dimension. This conclusion does not depend on any all-or-none assumptions about attention. Even if a subject does sometimes attend to the relevant cue during the early trials of discrimination training (and we have seen that it does), the probability of attending must still increase to some high value, in order that accurate performance should occur. Without being committed to an extreme noncontinuity theory, it remains possible to claim that part of what a subject must learn is the appropriate classification.

Is it possible then to provide evidence for the occurrence of such classificatory learning? Can one experimentally separate out two components in discrimination learning? Three classes of experiment have been devised which attempt to do just this: those on (*a*) the acquired distinctiveness of cues, (*b*) transfer along a continuum, and (*c*) reversal learning. In each case the credit for seeing their importance and for performing the first satisfactory experiment belongs to Lawrence. One of these areas (that of reversal learning) has been the center of an enormous volume of recent research, and an attempt will be made to resolve some apparent contradictions.

The Acquired Distinctiveness of Cues

If successful performance on a discrimination problem depends on two processes (learning to attend and learning choice responses), then speed of learning a given problem will depend upon the amount of learning of each that must occur (thus equally on how much of each has already occurred and will transfer to the new situation).

Ex hypothesi, animals cannot learn responses to stimuli unless they are also attending to those stimuli, but there is no reason why they should not learn to attend to a brightness cue, without learning the specific responses necessary for solving the normal simultaneous brightness discrimination. Thus if an animal is trained to attend to a particular stimulus dimension, by learning to make one type of response to stimuli varying along the dimension, and is then confronted with a second problem where the relevant dimension is the same but the type of response to be learned is different, solution of the second problem should be facilitated.

Lawrence's (1949) first experiment on the acquired distinctiveness of cues provides a direct test of this hypothesis: Subjects were trained on a simultaneous discrimination between A' and A'' (i.e., they were trained to approach A' and avoid A''); they then learned a successive discrimination (i.e., to turn right if one stimulus was present, and left if another was present). If the relevant stimuli of the successive discrimination were A'-A'', the subjects learned very much faster than if there were different relevant stimuli, or than if A'-A'' were irrelevant. However, simply learning the response of approaching A' and avoiding A'' could have been of no assistance in learning the successive discrimination. The only feature common to the two discriminations was that they shared the same relevant cue, and the positive transfer must have been due to the subjects' having learned to attend to this cue, and having transferred this learning between problems.

In Lawrence's (1950) second experiment, animals were trained on a successive discrimination, with A'-A'' the relevant dimension, and B'-B'' irrelevant. They then learned the combined simultaneous discrimination of $A'B'$ versus $A''B''$ and were finally given tests to determine the basis of their solution of the simultaneous discrimination. The tests established that the relevant successive cue was more effective than the irrelevant one.

It is clear, therefore, that irrespective of the responses required transfer can occur between discrimination problems on the basis of the relevant stimulus dimension. Learning to attend to a particular cue in one situation will transfer to another situation. Basically similar results have been obtained with rats by Mackintosh (1964) and Shepp and Eimas (1964); and with college students by, among others, Eckstrand and Wickens (1954).

Conversely, it ought to be possible to train animals to ignore a particular cue, and such training would be expected to retard the learning of a discrimination with that cue relevant. One possible way of training a subject not to attend to a particular dimension might be consistently to reward both values of that dimension shown to the animal, that is, to provide nondifferential reward on the dimension. (This possibility is implied by the suggestion put forward by Sutherland [1964] and by Wyckoff [1952] that what "switches in an analyzer" or "increases the probability of observing responses" is differential reinforcement of different stimuli.) If this is so, then such a procedure should retard subsequent discrimination learning, and results obtained by Siegel (1945) support this suggestion. However, later work by Bitterman, Calvin, and Elam (1953) and Bitterman and Elam (1954) seemed to show that the effect of nondifferential reward was even more harmful if given on a set of stimuli different from the test set. Although this effect has not been confirmed in other experiments (neither Billingsley, Feddersen & Bitterman, 1954, nor Calvin, 1953, found any difference), it does seem that if one wishes to claim that training to ignore a particular dimension will retard subsequent learning of that dimension, then one cannot also claim that the nondifferential reward technique is one way of training animals to ignore a dimension.

It remains possible that other ways of training animals to ignore a dimension might be more efficacious, and one obvious way would be to train animals with the dimension in question irrelevant (each stimulus equally often rewarded and not rewarded). An experiment by Goodwin and Lawrence (1955) showed fairly directly that subjects will stop attending to an irrelevant cue: Rats were trained on one discrimination, then on a second with the cue of the first present but irrelevant, and were finally retrained on the first problem either with the stimuli retaining their original signs, or with the stimuli reversed. Training on the second problem with the cue of the first present but irrelevant, should, if the subject continues to attend to that cue, tend to equalize the habit strength to each of the stimuli, and therefore eliminate any difference in the final stage between learning either the reversal or the nonreversal of the original problem. Goodwin and Lawrence found no such effect, and concluded that in the second stage subjects simply stopped attending to the original cue, and thus failed to equalize habit strength to the original S + and S −. Although this conclusion has been disputed by Goodrich, Ross, and Wagner (1961) and by Ross (1962, 1963), Mackintosh (1963c) has obtained results which support Goodwin and Lawrence's analysis.

There is independent evidence, therefore, that subjects trained with a cue irrelevant learn to ignore that cue; thus it should also be possible to demonstrate that training with a particular cue irrelevant retards subsequent learning with the cue relevant. There is evidence from Calvin's (1953) experiment mentioned above to support this: After receiving 16 days of nondifferential reward on horizontal-vertical striations, subjects learned the striation discrimination significantly faster than after receiving an equal number of nondifferential trials of which one in six was unrewarded (i.e., after pretraining with horizontal-vertical irrelevant). Jeeves and North (1956) claimed to be unable to find such an effect. They trained two groups on a horizontal-vertical discrimination, one with black-white irrelevant, one with no brightness cue present, and then trained them on the black-white discrimination. The first group made 15.08 errors, the second 12.92 errors during 80 test trials—a difference that fell short of significance. However, the difference was in the predicted direction, and it seems possible that it was underestimated, since the horizontal-vertical problem was actually learned faster by the group with black-white irrelevant. This is a very improbable result since Wortz and Bitterman (1953), for example, have shown that the presence of an irrelevant cue can significantly retard discrimination learning; thus it is probable that in Jeeves and North's experiment the group trained with the irrelevant cue contained faster learners, and, therefore, that had the two groups been matched in the first place, the irrelevant cue group would have been significantly inferior to the control group in the final test.

In summary, then, there is clear-cut evidence that pretraining animals to attend to a given dimension will benefit learning of a discrimination involving that dimension—irrespective of the relationship between the responses required in the two problems. There is some suggestion that the converse of this is also true: that it is possible to pretrain animals to ignore a particular dimension, and that such pretraining retards learning of a discrimination involving that dimension; but the evidence on this last point is by no means satisfactory, and there is need for further research.

Transfer along a Continuum

In the previous section, the argument was, in effect, that rate of learning should be influenced by the previously established probability of a subject's attending to the relevant cue. In this section, a similar argument is applied to accuracy of performance. For a subject to perform with a high degree of accuracy on any discrimination, the present hypothesis states that the subject must both have a high probability of attending to the relevant cue, and a high probability of selecting S + given that it is attending. Conversely, inaccurate performance might be due to either (or both) of these probabilities being substantially below unity.

Consider an animal trained on a discrimination between two neighboring shades of gray. From the close similarity of $S+$ and $S-$ two things are likely to follow: (*a*) There will be some overlap in the central representation of $S+$ and $S-$, so that even if the animal attends to the relevant cue on every trial, it will never attain 100% accuracy; (*b*) Precisely because mistakes will still be made when the animal is attending appropriately, it will not learn to attend consistently. Thus an animal performing at a low level of accuracy on a difficult discrimination does so both because it does not consistently attend to the relevant cue and because, even when it is attending, the stimuli are sometimes indistinguishable. Removal of either of these reasons, therefore, should improve accuracy.

Increased accuracy due to attending more consistently should be achieved by pretraining the subject on an easier discrimination in the same dimension until a high level of accuracy is attained, and then approaching the difficult discrimination by gradually decreasing the difference between the training stimuli. Early experiments by Volkmann on the two-point threshold (cited by James, 1890), by Pavlov (1927) on circle-ellipse discrimination by a dog, and by Lashley (1938) on a shape discrimination in rats, suggested that the predicted results do occur, but Lawrence's experiment (1952) provided the first controlled results on what he called "transfer along a continuum." The basic result has since been repeated by Baker and Osgood (1952) with human adults and a pitch discrimination, and by House and Zeaman (1960) with mentally retarded children, North (1959) with rats, and Sutherland, Mackintosh, and Mackintosh (1963) with octopuses, all using shape discriminations. Lawrence (1955) has noted the difficulties confronting orthodox continuity theory in attempting to explain the results.

It remains to show that in a difficult discrimination, accuracy also suffers from a subject's failure to distinguish between $S+$ and $S-$ even when attending appropriately. In the Sutherland et al. experiment already mentioned, subjects trained from the outset on the difficult discrimination were finally given transfer tests with the easier stimuli; without the opportunity for further learning, they showed immediate and significant increases in accuracy.

Reversal Learning

If an animal is trained to approach Stimulus A' and to avoid A'', and is then trained on the reversal of this problem, that is, to approach A'' and avoid A', then a single-stage continuity theory of discrimination learning predicts that prior A'-A'' training can only hinder A''-A' learning. For the type of two-stage model considered here, however, the position is more complicated, but for that very reason more interesting. For on this account, although the choice responses required in original learning and reversal are antagonistic, the relevant cue remains the same throughout, and therefore the attention established by original training is entirely appropriate for the reversal learning.

From this, it does not follow that reversal learning is bound to be very easy, let alone that rate of learning A''-A' should actually be increased by prior training on A'-A''. Although such a claim seems to be made by H. H. Kendler and T. S. Kendler (1962), it is surely fallacious. At the beginning of reversal an animal is fairly consistently unrewarded; on a two-stage model this will lead to extinction of both attention and choice responses. If the animal ceases to attend to the relevant cue before it has extinguished its original choice responses, then reversal learning can only be slow and laborious, since both attention and choice responses must be built up *de novo*. If, on the other hand, the original choice responses are extinguished while the subject is still attending to the relevant cue, then it has only to acquire newly appropriate choice responses, and the reversal is learned easily.

At first sight, it does not seem as if a two-stage model must necessarily predict one of these alternatives rather than the other; thus there is no need for it to predict that reversal learning should be relatively easy. It is clear, however, that one definite prediction can be made: If attention extinguishes before choice responses, reversal will be slower than if choice responses extinguish before attention, and any procedure that increases the probability of attending without equally increasing response strength, will facilitate reversal.

Once again it is Lawrence (1950) who

seems to have been the first to recognize the significance of reversal learning. He trained subjects on a successive discrimination with Cue *A* relevant and Cue *B* irrelevant, then on a simultaneous discrimination in which both were relevant, and finally on either *A* or *B* in isolation. In this last stage, half the subjects were trained with the same stimulus positive as had been positive in the previous stage, while the other half were trained on the reversal. As has already been noted, the nonreversal was learned faster on Cue *A* than on Cue *B*, implying that subjects had learned the combined discrimination largely in terms of Cue *A*. What was more surprising was that the reversal was also learned faster on Cue *A* than on Cue *B*. This result can be explained on the following set of assumptions: (*a*) that pretraining with *A* relevant and *B* irrelevant securely established attention to *A;* (*b*) that by the end of simultaneous discrimination training on the combined problem, subjects were attending to *A* with a very high probability and to *B* with a low probability; (*c*) that, consequently, during reversal to *A*, choice responses would extinguish faster than attention to the relevant cue while, during reversal to *B*, attention to the relevant cue would extinguish faster. This explanation may be lacking in rigor; at least, however, a two-stage theory seems capable of providing an explanation, and this is an achievement altogether beyond the scope of a single-stage model.

The Overtraining Reversal Effect. If pretraining with Cue *A* relevant and Cue *B* irrelevant is sufficient to ensure that attention to *A* is less likely to extinguish before choice responses than is attention to *B*, it seems plausible to argue that increasing the amount of training on *A* might have similar effects. Reid (1953) found that if rats are trained on a simultaneous brightness discrimination, given 0, 50, or 150 overtraining trials after reaching criterion, and are then reversed, speed of reversal learning is directly related to amount of overtraining—overtraining benefits reversal learning. Shortly afterwards the results were confirmed by Pubols (1956) and by Capaldi and Stevenson (1957).

Both Reid and Pubols attempted to explain their results in terms logically similar to those suggested here (although they employed the concept of the orienting or discriminating response instead of that of attention). Overtraining, they assumed, ensured that subjects would continue to look at or orientate toward the relevant stimuli throughout reversal whereas nonovertrained subjects would extinguish any tendency to look at the relevant stimuli and would "run blind." Stated in terms of attention, this explanation runs as follows: If overtraining increases the probability of subjects attending appropriately without causing a corresponding increase in choice response strength, then overtrained subjects will continue to attend to the relevant cue during reversal and will only have to extinguish their original choice responses and acquire new ones; while nonovertrained subjects will cease to attend to the relevant cue, attend to other irrelevant cues (during which time they will learn little about the correlation of the relevant cue with reward), and hence will take a long time to learn the reversal. This explanation of the overtraining reversal effect was first suggested by Sutherland (1959) and has since been stated more formally by Lovejoy (1965).

At least two other major categories of explanation have been advanced for the overtraining reversal effect. Capaldi and Stevenson (1957) suggested that overtraining might benefit reversal by reducing resistance to extinction. Such a suggestion could supposedly follow either from the discrimination hypothesis of extinction (advanced by Capaldi and Stevenson, 1957, and also by Guttman, 1963), or from a motivational (frustration) account of extinction (Amsel, 1962; Birch, 1961), or for yet other reasons (D'Amato & Jagoda, 1960, 1961).

Alternatively, overtraining might benefit reversal by reducing differential response tendencies to irrelevant cues (specifically position). This suggestion was first put forward by Spence (cited by Reid, 1953) and has since been propounded rather more generally by Harlow (1959). According to Spence, this effect would occur because any bias towards selecting one position rather than another is due to a difference in the excitatory tendency elicited by each position, and increasing the number of trials on which each position is equally often rewarded and not rewarded would gradually equate these excitatory values.

It should be noticed that neither of these explanations of the effect of overtraining on reversal is able to account for Lawrence's (1950) reversal results. In Lawrence's experiment the distinction between the two groups lay not in the amount of training that each had received but in its relevance to the reversal task, and this notion of relevance seemed to lie entirely outside the scope of these two suggestions. This does not necessarily imply that the explanation they provided of the overtraining reversal effect is inadequate, only that the attention theory can explain a wider range of results. In fact, however, there is overwhelming evidence against the extinction account of the overtraining reversal effect, and quite adequate evidence against the irrelevant response adaptation account.

The major objection to an explanation of the overtraining reversal effect, in terms of faster extinction following overtraining, is that such an account does not even provide an accurate description of reversal learning data. Empirically it is presumably justifiable to regard reversal learning as comprising two parts: (*a*) extinction of the tendency to select the former S +, and (*b*) acquisition of a tendency to select the new S +. Stage (*a*) is the extinction phase and might be regarded as continuing as long as the subject scores less than 50% correct, while Stage (*b*) starts as soon as the subject scores 50% or better. At a descriptive level, therefore, the extinction account implies that overtraining facilitates reversal by shortening Stage (*a*). There is no evidence whatsoever to support this claim. Confusion has arisen because of the undoubted fact that in a runway increased training does often reduce resistance to extinction (Hill & Spear, 1963a; Ison, 1962; Ison & Cook, 1964; Madison, 1964; North & Stimmel, 1960; Siegel & Wagner, 1963; Wagner, 1963). Thus in the experiment of Birch, Ison, and Sperling (1960), where discrimination was defined in terms of running speed in a runway, the overtraining reversal effect obtained can be economically explained in terms of extinction. But this does not necessarily imply that overtraining in a choice situation will reduce resistance to extinction of choice responses, and in fact it does not; it increases resistance to extinction. All investigators of reversal learning who have actually published data on their subjects' tendency to continue selecting their former S + during the first half of reversal learning (whatever precise measure of this tendency they have used) have uniformly found that overtraining increases resistance to extinction (Komaki, 1961; Mackintosh, 1962, 1963b, 1963c, 1965b; Reid, 1953). Overtraining facilitates reversal of a simultaneous visual discrimination not because of, but in spite of its effect on extinction.

It is true that in several of these cases, the measure of choice response extinction was the number of initial consecutive errors at the outset of reversal, and that one might argue (as does, for example, Sperling, 1965) that greater initial resistance to extinction following overtraining might be followed by subsequently more rapid extinction. However, Mackintosh (1963c) showed that overtraining increased the number of trials taken to reach an extinction criterion of equal choice of S + and S − over 10 trials. In the light of this evidence, any further suggestion that overtraining facilitates reversal by producing faster extinction must be using the term "extinction" in a way quite different from that in which it is used here.

Much of the evidence to be discussed below is also incompatible with the extinction account of the overtraining reversal effect; for example, experiments on the effects of irrelevant cues on reversal learning and on nonreversal shifts produce results either inexplicable in terms of this account or actually contradicting it. Nor does it seem easy to account for the basic difference between the results of visual and spatial reversal experiments in this way (but see Sperling, 1965). These arguments, however, are not to be taken as denying (what must be true) that, other things being equal, rate of extinction and rate of reversal will be positively correlated. What is being denied is that in simultaneous visual discriminations other things are equal. Certainly, as Theios and Blosser (1965) have claimed, it is possible that reward magnitude is an important variable in determining whether overtraining facilitates reversal, because reward magnitude and overtraining interact in their effect on extinction. But this would not constitute an explanation of the overtraining reversal effect solely in terms of resistance to extinction. It might simply be that with small rewards overtraining so greatly increases resistance to extinction that subsequent savings are outweighed, while with larger rewards overtraining only slightly increases resistance to extinction, so that subsequent savings are sufficient to produce overall faster reversal learning.

Remaining at a descriptive level, all investigators who have published the evidence (the same as those who have reported evidence on extinction) have agreed in showing that overtraining facilitates reversal of a simultaneous brightness discrimination because it shortens the run of position responses during the middle of reversal. This fact would seem to support the second alternative—the suggestion that overtraining facilitates reversal by reducing the tendency to respond to irrelevant cues—but it is possible that the cause of this reduction in responses to irrelevant cues is not that envisaged by Spence, but is, on the contrary, due to an increase in the tendency to respond to the relevant cue, that is, an increase in the probability of attending.

Although both explanations predict that the effect of extra irrelevant cues will be to increase the magnitude of the overtraining reversal effect (a prediction confirmed for rats by Mackintosh, 1963b; for chicks by Mackintosh, 1965b; for octopuses by Mackintosh & Mackintosh, 1963), and although both predict (correctly) that this effect appears because extra irrelevant cues seriously retard reversal for nonovertrained subjects but not for overtrained subjects, Spence's account (cited by Reid, 1953) claims that this is because overtrained subjects have had further opportunity to "adapt out" the irrelevant cues, while the attention theory claims that it is because they have had further opportunity to learn to attend to the relevant cue. Decision between the two accounts can be reached by delaying the introduction of the extra irrelevant cue until reversal. Overtrained animals, although they have had no opportunity to adapt out the cue, are still able to ignore it during reversal (Mackintosh, 1963b), and this can only be because they have learned to attend to the relevant cue.

More decisive evidence against Spence's explanation of reversal learning comes from experiments on nonreversal shifts to be considered below. Further evidence in favor of attention theory comes from experiments which have shown that the relative rates of extinction of attention and choice responses do indeed vary as a function of overtraining in the way predicted by the theory. To recapitulate: The attention explanation of the overtraining reversal effect is based upon the assumption that overtraining increases the probability of attending to the relevant cue during reversal without causing a corresponding increase in choice response strength. The result of this is that overtrained subjects extinguish their choice responses while still attending to the relevant cue, whereas nonovertrained subjects extinguish their tendency to attend to the relevant cue before fully extinguishing choice responses. A direct test of this hypothesis was to train subjects on a given discrimination with or without overtraining and then to extinguish them until they selected S+ and S− equally often. When reversed, overtrained subjects, who had merely extinguished choice responses only, had merely to establish newly appropriate responses, while nonovertrained subjects, who had ceased attending to the relevant cue without having extinguished choice responses, when they attended to the relevant cue again, started responding in the direction in which they were originally trained. In two experiments with rats (Mackintosh, 1963c, 1965b), this difference was demonstrated: Upon reversal after extinction, overtrained subjects showed an orderly increase in choice of their new S+, but nonovertrained subjects (as was first shown by D'Amato and Jagoda, 1960) showed an initial increase in choice of their former S+.

Variability of the Effect. A number of investigators have claimed that the overtraining reversal effect is an elusive, even suspiciously inconsistent, phenomenon. Doubtless there is much variability in the results of psychological experiments owing to variations in experimental techniques, but there is no evidence that the overtraining reversal effect is unusually prone to inexplicable disappearances and sudden reappearances. Any suggestion that this is so seems to be partly due to confusion of two separate classes of experiment—reversal of visual discriminations and reversal of spatial discriminations. So far we have been talking about reversal of simultaneous visual (usually brightness) discriminations. To date there have been 13 such experiments on rats, the results of which are presented in Table 1.

It has also been possible to obtain the effect in octopuses (Mackintosh & Mackintosh, 1963), chicks (Mackintosh, 1965b), and children (Youniss & Furth, 1964a, b), provided the discrimination was relatively difficult and contained irrelevant cues. In spite of Erlebacher's (1963) single failure, and even the repeated failure of D'Amato and Schiff (1965) to obtain the effect, the results of these experiments are not particularly contradictory. It can certainly not be claimed that the effect is an elusive one.

On the other hand, the results of position reversal experiments (shown in Table 2), although somewhat more conflicting, are nearly as decisive in failing to find the effect. This makes a total of 22 experiments, in only seven of which did overtraining facilitate reversal. The result in the "no effect" column was usually that overtraining retarded reversal, but not significantly.

The problem, then, is how to explain both the general difference between the visual and the spatial discrimination experiments, and the conflicting results within each class of study. One fairly obvious suggestion is that visual problems contain irrelevant cues, and spatial problems in general do not. It has of course already been shown that irrelevant cues are important in producing the effect. In support of this suggestion is the fact that two of the spatial experiments in which overtraining facilitated reversal (Brookshire, Warren, & Ball, 1961, and Pubols, 1956) did actually involve irrelevant visual cues, while a third (Bruner et al., 1958) was a relatively complex alternation learning problem in a 4-unit maze, instead of the normal simple T-maze problem. Similarly, Youniss and Furth (1946b) found that overtraining only facilitated reversal of a spatial discrimination in children when there were irrelevant cues. Furthermore, even in a brightness discrimination the experimental situation may be such as to minimize the importance of irrelevant cues, and thereby reduce the chances of finding an effect. It is possible that the automatic apparatus used by D'Amato and Schiff (1965) in their brightness reversal

TABLE 1

RESULTS OF VISUAL DISCRIMINATION REVERSAL EXPERIMENTS WITH RATS

Overtraining facilitates reversal	Overtraining has no effect on reversal
Brookshire, Warren, and Ball (1961)	Erlebacher (1963)
Capaldi and Stevenson (1957)	D'Amato and Schiff (1965)
D'Amato and Jagoda (1961)	
Komaki (1961)	
Mackintosh (1962, 1963b, 1965b)	
North and Clayton (1959)	
Pubols (1956)	
Reid (1953)	
Sasaki (1960)	

TABLE 2

RESULTS OF SPATIAL DISCRIMINATION REVERSAL EXPERIMENTS WITH RATS

Overtraining facilitates reversal	Overtraining has no effect on reversal	Overtraining retards reversal
Brookshire, Warren, and Ball (1961)	Clayton (1963a, 1963b)	Hill and Spear (1963b)
Bruner, Mandler, O'Dowd, and Wallach (1958)	D'Amato and Jagoda (1962)	Hill, Spear, and Clayton (1962)
Capaldi (1963)	D'Amato and Schiff (1964)	Krechevsky and Honzik (1932)
Ison and Birch (1961)	Galanter and Bush (1959)	Mackintosh (1965c)
Pubols (1956)	Hill et al. (1962)	
Takemura (1961)	Iwahara (1959)	
Theios and Blosser (1965)	Komaki (1962)	
	Mackintosh (1965c)	
	Paul and Kesner (1963)	
	Theios and Blosser (1965)	

experiments provided very few irrelevant spatial cues—certainly rats seemed to take a long time to learn a simple position habit in this apparatus (D'Amato & Jagoda, 1962). On the other hand, as pointed out by Lovejoy (1965), Pubols' results are not reliable, since his two groups were not properly matched in initial learning speed, and Clayton (1963b) performed a direct test of the effect of irrelevant cues on spatial reversal with negative results (in fact, he claimed that irrelevant cues actually retard overtrained subjects more than nonovertrained subjects, but his claim is unconvincing since he failed to find any difference in overall trials to criterion).

The question of the importance of irrelevant cues in spatial reversal is one that deserves further investigation. Nevertheless, it is extremely unlikely that this is the only critical difference between visual and spatial problems. Mackintosh (1965c) has suggested that one reason why overtraining does not benefit spatial reversal is that rats given increasing amounts of training on a maze habit become increasingly likely to run off internally controlled response sequences independent of external cues. Direct evidence for this proposition was provided by the finding that overtraining in a T maze can turn rats from place learners into response learners (Hicks, 1964; Mackintosh, 1965c). Again this would not be the only relevant factor, but it not only serves to explain the difference between visual and spatial experiments, it also helps to explain differences within the spatial studies (see Mackintosh, 1965c).

A third possible explanation in terms of an attention theory has been suggested by Lovejoy (1965). He points out that the overtraining reversal effect is assumed to depend on nonovertrained subjects ceasing to attend to the revelant cue during reversal. But if the relevant cue is one that subjects in this situation automatically attend to, then presumably nonovertrained subjects will continue to do so throughout reversal. Rats learn spatial discriminations very fast and are generally reckoned to be dominantly spatially oriented; therefore, spatial cues have a high priority for rats, and are automatically attended to. Under these circumstances, a two-stage theory contracts into a single-stage theory and will not predict an overtraining reversal effect. Evidence bearing on this general point of view has been provided by Mackintosh (1965b). Chicks are dominantly visual animals, learn simple visual discriminations very fast, and do not show the overtraining reversal effect unless provided with an extremely difficult discrimination.

Two other possibilities should be briefly mentioned. Theios and Blosser (1965) have argued that the overtraining reversal effect only appears when a large reward is used. The basis of this suggestion is the assumption that reward magnitude and overtraining interact to determine resistance to extinction, an assumption that receives support from the study of Ison and Cook (1964), but not from that of Madison (1964). This can at best be only a partial explanation, since variations in the overtraining reversal effect can occur in the absence of changes in reward magnitude (e.g., Mackintosh, 1963b, 1965b); it may nevertheless have some validity. Finally it is worth noting that D'Amato and Schiff (1965) in all but one of their experiments gave 240 overtraining trials. There is in fact no evidence from any published study that overtraining to this extent will ever facilitate reversal learning.

It can be seen that there is no shortage of possible explanations for the differences between various experimental results, and that several of these explanations stem from an attention theory. It should be emphasized that these explanations are not mutually exclusive. There is no reason to suppose that one is right and all the others wrong, and every reason to believe that the overtraining reversal effect, like many other psychological phenomena, is overdetermined.

Before leaving the research on reversal learning, one final point must be made. The overtraining reversal effect is no doubt a rather dramatic and even paradoxical finding, and it may be for this reason that research has been concentrated almost exclusively on overall trials or errors to criterion scores. There have been relatively few attempts to examine in detail possible differences in the course of reversal learning between overtrained and nonovertrained subjects. Yet these are the critical differences which any explanation of the overtraining reversal effect must handle, and which must in fact decide between alternative explanations. For example, the occasional failure to find an overtraining reversal effect in brightness discrimination problems might be either because the irrelevant spatial cue was not a dominant one, or because some factor was increasing resistance to extinction in overtrained subjects. A casual inspection of the daily reversal scores would decide between these alternatives, but such scores are virtually never provided. In the future, experiments which fail to provide the evidence necessary for such decisions have little hope of advancing our knowledge. There is no need for a proliferation of further studies giving bare trials to criterion data.

Reversal and Nonreversal Shifts. H. H. Kendler and T. S. Kendler (1962) and their associates have laid almost exclusive emphasis on the importance of a comparison between rates of learning reversal and nonreversal shifts as evidence for or against a two-stage theory. I have already argued that this emphasis is misplaced, since the predictions a two-stage theory makes about speed of reversal learning depend on further specification of the relative rates of extinction of attention and choice responses. Nevertheless, it is true, as they recognize, that it is possible to explain the superiority of reversal to nonreversal shifts in terms of a two-stage theory, whereas (once partial reinforcement has been equated) it does not seem possible to do so in terms of a single-stage theory.

Evidence that reversals are solved faster than nonreversal shifts comes from experiments with college students (Buss, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955), and children over the age of five (Kendler & Kendler, 1959). Younger children (Kendler, Kendler, & Wells, 1960) and rats (Kelleher, 1956) learn nonreversals faster than reversals. From this it is claimed that young children and rats are "one-stage learners." But if it is not a necessary consequence of two-stage theory that reversals should be learned faster than nonreversals, there is nothing embarrassing about finding the opposite. In fact, more careful consideration shows that, in view of the explanation of the overtraining reversal effect just provided, this is precisely what would be expected. If animals are trained to criterion, they will only learn a reversal faster than a nonreversal if criterion training is sufficient to guarantee that attention extinguishes more slowly than choice responses. But this is exactly what has been assumed not to happen; in rats this is assumed only to occur after overtraining.

In rats, therefore, criterion training should produce faster learning of nonreversals than of reversals, while overtraining should at the least reduce this difference. This is true for young children—compare the results of Kendler, Kendler, and Wells (1960) with those of House and Zeaman (1962)—and although no one experiment has provided a direct test of the proposition in rats, the evidence is strongly in its favor. Kelleher (1956) has confirmed the first part, while Mackintosh (1962) has shown that although overtraining benefits reversal learning it retards nonreversal shift learning.

This last result has been confirmed on several occasions (in rats by Mackintosh, 1963c, 1964; in octopuses by Mackintosh & Mackintosh, 1964). However, Brookshire et al. (1961) and Uhl (1964) failed to find any effect, and Komaki (1961) has claimed that overtraining benefits nonreversal learning. A possible source of confusion here is that both Brookshire et al. (see Mackintosh, 1964) and, in one of his experiments, Komaki (see Law, 1954; Mackintosh, 1963a) may have used a nonreversal discrimination that involved the same relevant cue as did original learning. If overtraining increases the probability of a subject attending to a particular stimulus dimension, then this would account for their results. Uhl's results are not strictly comparable with other experiments, since he trained his subjects on a partial reinforcement schedule. Sutherland's (1964) explanation of the effects of partial reinforcement in terms of an attention theory would, in fact, probably predict this result of Uhl.

If it can be accepted that, at the least, overtraining does not facilitate nonreversal learning, then this decisively rules out any explanation of the overtraining reversal effect that fails to take into account the fact that a reversal involves the same relevant cue as original learning. It is obvious, for example, that a simultaneous brightness discrimination and a simultaneous orientation discrimination share irrelevant (position) cues in common, and if overtraining is supposed directly to reduce the probability of responding to irrelevant cues, it must benefit nonreversal learning at least as much as, and possibly more than, it benefits reversal.

THEORETICAL INTERPRETATIONS

This survey of the relevant experimental evidence has established two main claims: In solving discrimination problems, animals (*a*) do not attend equally to all features of the stimulus situation, and (*b*) must learn which features to attend to. One of the major theses of this review is that there is a logical connection between these two claims, since if an animal automatically classified its input in all ways simultaneously, there would be no need for it to learn to perform the appropriate classification. Such a need only arises if animals do not automatically attend to the relevant stimulus dimension.

Surprisingly, the two positions seem to have grown up more or less independently. Noncontinuity theorists like Lashley and Krechevsky introduced the concept of attention into discussions of discrimination learning by claiming that animals attended to only a single cue at a time, but they made little systematic attempt to consider what implications this position would hold for the analysis of the learning process. Conversely, there has recently been a great growth of interest among relatively orthodox S-R theorists in various two-stage models of discrimination learning, but very often the connection with the original question of attention has been totally ignored. Among other possibilities, it has been suggested that the subject must acquire relevant mediating responses (Kendler & Kendler, 1962), or observing responses (Miller & Dollard, 1941; Wyckoff, 1952), or orienting responses (Spence, 1960). Rather closer to the spirit of this review, others have suggested that the subject must attend to the relevant dimension (Goodwin & Lawrence, 1955), learn to switch in the relevant analyzer (Sutherland, 1959), or learn attention responses (Lovejoy, 1965; Zeaman & House, 1963), while Lawrence (1949, 1963, respectively) has also talked of the acquired distinctiveness of cues, and of stimulus coding.

Although there are clear logical similarities between all these theoretical suggestions, there are also differences of substance between them, and in the following account these differences may well appear to be exaggerated, but such exaggeration is not merely perverse. In the light of the evidence there can be no doubt that some sort of two-stage account of animal discrimination learning must be provided, and up to this point there may well be agreement. But if theoretical progress is to be made, it must be partly along the lines of specifying more precisely the exact nature of the two stages. This section represents a preliminary attempt to classify the models proposed so far, and very briefly considers some of the evidence that will be relevant to deciding between them.

For expository purposes, the models can be divided into three classes: (*a*) orienting or observing response models (Spence, Wyckoff), (*b*) mediate responses models (Kendler & Kendler), and (*c*) attention models (Lawrence, Sutherland, Zeaman & House).

Orienting Response Theories

To S-R theorists there seems to have been something doubly distasteful about the concept of attention. In a very general sort of way, its use would appear to imply a reversion to a mentalistic, even introspective psychology; more specifically, it would appear to imply a central mechanism, and S-R theory is happiest when dealing with external responses. The concept of the orienting response, in fact, has been put forward as an external substitute for attention.

Already in the earliest formulation of his theory of discrimination learning, Spence (1936) noted that "the mere presence of the cue stimulus somewhere in the experimental situation does not guarantee its impingement on the animal's sensorium at or near the critical moment of response [p. 438]." He later elaborated this by saying that solution of a discrimination problem involves the learning of

> responses which lead to the reception of the appropriate aspects of the total environmental complex on the animal's sensorium, e.g., the orientation and fixation of head and eyes towards the critical stimuli. That is, the animal learns to "look at" one aspect of the situation rather than another [1937, p. 432].

Thus, when confronted with evidence that animals had failed to attend to the relevant cue in Krechevsky's (1938) presolution period experiment, Spence (1940) claimed that the animals "did not see the rows of black squares because they were fixating other aspects of the stimulus complex [p. 281]."

As Lashley (1942) insisted, this particular claim was somewhat implausible from purely physical considerations of the angular separation of different parts of the stimulus cards in a jumping stand, and of the rat's field of vision; and as Goodrich et al. (1961) have more recently pointed out, other criteria of when rats are looking at a particular stimulus turn out dangerously circular. In spite of these difficulties, and in spite of the almost total lack of evidence showing the importance of overt orienting responses in discrimination learning by rats—Ehrenfreund's results (1948) were seemingly contradicted by Mahut (1954)—this type of account has become increasingly popular—Spence's most recent statement of his theory laid considerable stress on orienting responses (Spence, 1960), and both Pubols (1956) and Reid (1953) attempted to explain the overtraining reversal effect in these terms.

This popularity is partly due to the ingenious work pioneered by Wyckoff (1952), where animals literally had to learn to make a specified response in order to see the discriminative stimuli. The ingenuity of such experiments and the interest of their results should blind us neither to the fact that they do not manipulate the orientation of an animal's receptors at all, nor to the possibility that in normal discrimination learning experiments no such receptor orientation is necessary, since in any suitable apparatus the subject automatically sees the relevant stimuli.

Fortunately it is possible to go further than this and to show experimentally that attention cannot be adequately conceptualized in terms of orienting responses. Goodwin and Lawrence (1955) showed selective attention operating in situations where the relevant stimuli were apparently "perceptually obvious" and therefore, they argued, on Spence's account, discriminable without elaborate orienting responses. Their argument is plausible, but not compelling. But if rats are trained on a discrimination between a black horizontal and a white vertical rectangle, there is no conceivable way in which they could orientate themselves so as to look at the orientation differences between the stimuli without thereby seeing the brightness difference (and vice versa). Thus, demonstration of selective attention between these dimensions in this situation decisively rules out any orienting response interpretation (Mackintosh, 1965a).

The writer has also specifically tested the explanation of the effect of overtraining in terms of orienting responses proposed by Pubols (1956) and Reid (1953). If overtrained animals are more likely to look at the appropriate physical stimuli, then overtraining on a square versus diamond discrimination, in which both stimuli are sometimes black and sometimes white, could only benefit the learning of a black- versus white-square discrimination. In fact, as is predicted by attention theory, overtraining retards learning in this situation (Mackintosh, 1964). Thus overtraining benefits reversal learning, not because the physical stimuli of the reversal problem were present during original training, but because both problems involve the same stimulus *dimension*. It is just not reasonable to suggest that any orientation of an animal's receptors could be responsible for selection between dimensions of stimuli.

Mediating Response Theories

The notion of the orienting response shares one feature in common with the concept of attention: At least it is clear that if animals fail to establish appropriate orientation of their head and receptors they will fail to solve the discrimination problem. However, it is also clear that there is a sense in which animals attend to some cues and ignore others, which cannot be characterized by talking of overt orienting responses. Perhaps some such considerations have led other S-R theorists into adopting a mediating response theory, but there has been a tendency to leave the precise nature of these mediating responses unspecified, and consequently no such arguments as those above can be marshalled against the theory.

This vagueness, however, generates some further difficulties, of which the foremost is that the function of these mediating responses remains very unclear. Kendler and Kendler (1962), for example, do not regard the occurrence of mediating responses as necessary for discrimination learning, since children under the age of five and all nonhuman animals are assumed not to have such responses available. There is no evidence that older children and adults are notably superior at solving simple discrimination problems to these subjects (e.g., Stevenson, Iscoe, & McConnell, 1955), so that it is not obvious what advantage the older humans are supposed to gain from the possession of mediating responses. Unless the mediating responses in some way assist learning, it is hard to see how they themselves become established. Doubtless one mediating response might receive more regular reinforcement than another, but there seem no grounds for supposing that either would be more regularly reinforced than no mediating response would be.

If in this respect the concept of the mediating response seems altogether too vague, there is another respect in which it has been too strictly circumscribed to remain capable of explaining the evidence. It is, supposedly, a truism that S-R theory has incorporated into itself all features of value from more cognitive theories; just as expectancies become fractional anticipatory goal responses, so attention becomes a mediating response. One justification for this translation is said to be that we can now apply the laws obtained with overt S-R connections to describe the behavior of these internal or inferred responses. Kendler and Kendler (1962) wrote:

> The basic assumption of the mediational hypothesis, at least for the time being, is that the implicit stimulus and response events obey the same principles that operate in observable S-R relationships [p. 6].

This assumption is not further discussed; but if held strictly, it invalidates the explanation of their results. In order to explain the superiority of reversal to nonreversal shift learning, Kendler and Kendler claimed that:

> According to a single unit hypothesis . . . a nonreversal shift . . . would occur more rapidly than a reversal shift. The reason for this is that at the time of the shift the difference between the strength of the dominant incorrect habit and the to-be-correct habit is much greater for the reversal, as compared to the nonreversal shift. According to the mediational theory the situation is entirely different. A reversal enables the subject to utilize the same mediated response. Only the overt response has to be changed. A nonreversal shift, on the other hand, requires the acquisition of a *new* mediated response, the cues of which have to be attached to *new* overt responses. Because the old mediational sequence has to be discarded and a new one formed, the nonreversal shift should be executed more slowly than a reversal shift [p. 7].

The flaw in this argument has already been pointed out in an earlier section. A reversal shift will only enable "the subject to utilize the same mediated response," if the mediated response is still retained after the original overt responses are extinguished. If both classes of response are extinguished at the same rate, then by the time the overt responses are extinguished, there will be no mediated response available for connection to the new overt responses. The subject will have to reestablish the old mediating response in order to learn a reversal, just as he will have to establish a new mediating response in order to learn a nonreversal shift. Thus on the assumption of equal rates of extinction, there are no grounds for deciding which type of problem will be learned more easily. On the other hand, if the mediating response extinguishes faster than the overt responses, it will presumably be easier to learn a nonreversal than a reversal shift. Kendler and Kendler, in a footnote to the passage just quoted, recognized that their account was incomplete and suggested two ways of supplementing it. Either they assumed, as here, that mediating responses extinguish more slowly than overt responses, or, following Goss (1961), they claimed that one can "simply count the number of new associations that have to be formed" in order to learn reversal or nonreversal shifts. They suggested that more must be formed for nonreversal than for reversal learning. But this is still only true on certain assumptions about the relative rates of extinction of mediating and overt responses. If, for example, the mediating response extinguishes faster, then to learn a nonreversal the subject simply extinguishes the old mediating response, establishes a new one, and attaches new overt responses (making a total of three connections); while to learn a reversal, the subject must extinguish the old mediating response, extinguish its old overt responses, reestablish the old mediating response, and establish new overt responses (making a total of four connections).

It is likely, then, that in order to explain in terms of this type of theory the greater ease of reversal to nonreversal shift learning, the mediating response must be assumed to extinguish more slowly than the overt response. But if a two-stage model is constructed, and if it incorporates equivalent assumptions about the laws governing the two stages (including equal values for the paramters), then it turns out that precisely the opposite effect is generated (see Lovejoy, 1965). During learning the correct overt response increases at a faster rate than the correct mediating response, so that if training is taken to criterion only, the overt response has the greater strength; as a consequence of this it is the mediating response that reaches any given stage of extinction first.

If, therefore, the claim that "implicit stimulus and response events obey the same principles that operate in observable S-R relationships" is taken seriously, a mediating response theory appears to yield the same prediction as a single-stage theory, namely, that a nonreversal should be easier to learn than a reversal shift. At the very least Kendler and Kendler must discard this assumption in order to explain their own results. A more general complaint against the assumption of equivalence of laws governing mediating and overt responses is that it imposes severe restrictions on the range of data that can be handled by a two-stage theory. If the assumption is discarded, it becomes possible to explain apparently qualitative differences in the behavior of different animals in reversal learning situations by making detailed quantitative changes in the values of the various parameters. Both the differences between rats and college students in reversal versus nonreversal shift learning and the difference between rats and fish in the effect of overtraining on reversal learning should be explicable in these simple terms, and this exciting possibility of actually explaining species differences in learning experiments seems to be denied by the assumptions of S-R theories in this area.

Attention Theories

The criticisms made above of orienting and mediating response theories can be summarized by listing three requirements of a more adequate two-stage theory.

1. The attention stage should serve the function of permitting the animal to discriminate a given stimulus dimension. Evidence for this comes from all experiments demonstrating that animals do not learn about all cues equally.
2. This effect cannot be exclusively due to overt orientation of the animal's receptors. Attention selects between stimulus dimensions both in humans (Harrow, 1964; Isaacs and Duncan, 1962), and in rats (Shepp and Eimas, 1964), and not between stimulus objects (Mackintosh 1964) as is implied by orienting response theory.
3. There is no evidence to support the assumption that attention and choice responses obey exactly the same laws, and the possibility of altering the parameters of one independently of the other generates explanation of a wide range of comparative data.

Although, therefore, the similarities between different two-stage theories are considerable, there remain important differences between them. It is possible to test some of these differences experimentally, and the results of such experiments appear to lend support to the versions proposed by Goodwin and Lawrence (1955), Sutherland (1959), and Zeaman and House (1963).

CONCLUSIONS

The available evidence both supports a two-stage model of discrimination learning as opposed to a single-stage model, and within the class of two-stage models supports some of the implications of an attention theory as opposed to other theories. At the least, then, it would seem worthwhile to explore the potentialities of attention theories of learning further. Two possibilities suggest themselves.

On the one hand, if the concept of attention is one that can be profitably applied in the area of discrimination learning, it is worth considering its application to other problems. It is tempting, for example, to suggest that the absence of latent learning under some conditions, such as high irrelevant drive (Johnson, 1952) or prior overtraining (Bradley & George, 1962), can be explained in terms of a failure of attention to the relevant goal object; or that some of the differences between ordinary conditioning and sensory preconditioning (Hoffeld, Kendall, Thompson, & Brogden, 1960) reflect not a basic difference in learning processes but a difference in the extent to which an animal will attend to a bell and a light repetitively paired and a bell and food repetitively paired. These, of course, are only rather vague suggestions. As a more specific example, Sutherland (1964) has argued that the partial reinforcement effect can be understood in terms of an attention theory, and some preliminary evidence (Sutherland, Mackintosh, & Wolfe, 1965) supports this interpretation.

The second line of development must be the production of not only more general but also more formal theories. This is not simply because the concept of attention is a mentalistic one and needs to be "mechanized" in order to attain scientific respectability (models put forward by both Broadbent, 1957, and Sutherland, 1959, are cast in a mechanical form). The objection to many of the theories currently available is not so much to their language as to their lack of rigor. With informal theories, predictions tend to be derived on a partly intuitive basis, and it is often extremely difficult to decide whether a particular prediction does logically follow from a particular theory. To take one apt example, much of the argument in the previous section was concerned with whether Kendler and Kendler's mediating response theory did necessarily predict that a reversal would be easier to learn than a nonreversal shift. But if the theory were sufficiently precise there would be no room for verbal argument; either the question would be soluble analytically, or the model could be used to program a computer, and "stat-subjects" could be run in the appropriate experimental situation. Either procedure would give an unambiguous answer to the question.

Fortunately some preliminary steps have already been taken. The mathematical theory of Zeaman and House (1963) has been successfully applied to the behavior of retarded children in a number of situations. And Lovejoy (1965) has proposed a mathematical model which he has shown will predict the overtraining reversal effect. Exploration of the potentialities of these theories would seem to promise considerable rewards.

REFERENCES

Amsel, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review*, 1962, **69**, 306–328.

Baker, R. A., & Osgood, S. W. Discrimination transfer along a pitch continuum. *Journal of Experimental Psychology*, 1954, **48**, 241–246.

Billingsley, B. A., Feddersen, W. E., & Bitterman, M. E. Discrimination following nondifferential reinforcement with differential afferent consequences. *American Journal of Psychology*, 1954, **67**, 335–337.

Birch, D. A motivational interpretation of extinction. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1961*. Lincoln: University of Nebraska Press, 1961. Pp. 179–202.

Birch, D., Ison, J. R., & Sperling, Sally E. Reversal learning under single stimulus presentation. *Journal of Experimental Psychology*, 1960, **60**, 36–40.

Bitterman, M. E., Calvin, A. D., & Elam, C. B. Perceptual differentiation in the course of nondifferential reinforcement. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 393–397.

Bitterman, M. E., & Coate, W. B. Some new experiments on the nature of discrimination learning in the rat. *Journal of Comparative and Physiological Psychology*, 1950, **43**, 198–210.

Bitterman, M. E., & Elam, C. B. Discrimination following varying amounts of nondifferential reinforcement. *American Journal of Psychology*, 1954, **67**, 133–137.

Blum, R. A., & Blum, J. S. Factual issues in the "continuity controversy." *Psychological Review*, 1949, **56**, 33–50.

Bower, G., & Trabasso, T. Reversals prior to solution in concept identification. *Journal of Experimental Psychology*, 1963, **66**, 409–418.

Bradley, J. I. Pre-solution trials in discrimination learning. *Psychological Reports*, 1961, **8**, 155–158.

Bradley, J. I., & George, R. J. Conditions influencing latent learning. *Psychological Reports*, 1962, **11**, 499–505.

Broadbent, D. E. A mechanical model for human attention and immediate memory. *Psychological Review*, 1957, **64**, 205–215.

Broadbent, D. E. *Perception and communication.* London: Pergamon Press, 1958.

Broadbent, D. E. Human perception and animal learning. In W. H. Thorpe and O. L. Zangwill (Eds.), *Current problems in animal behaviour.* Cambridge: University Press, 1961. Pp. 248–272.

Brookshire, K. H., Warren, J. M., & Ball, G. G. Reversal and transfer learning following overtraining in rat and chicken. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 98–102.

Bruner, J. S., Mandler, J. M., O'Dowd, D., & Wallach, M. A. The role of overlearning and drive level in reversal learning. *Journal of Comparative and Physiological Psychology,* 1958, **51**, 607–613.

Bruner, J. S., Matter, J., & Papanek, M. L. Breadth of learning as a function of drive level and mechanization. *Psychological Review,* 1955, **62**, 1–10.

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *Journal of Experimental Psychology,* 1956, **52**, 152–166.

Butter, C. M. Stimulus generalization along one and two dimensions in the pigeon. *Journal of Experimental Psychology,* 1963, **65**, 339–346.

Calvin, A. D. The growth of learning during nondifferential reinforcement. *Journal of Experimental Psychology,* 1953, **46**, 248–254.

Capaldi, E. J. Overlearning reversal effect in a spatial discrimination task. *Perceptual and Motor Skills,* 1963, **16**, 335–336.

Capaldi, E. J. & Stevenson, H. W. Response reversal following different amounts of training. *Journal of Comparative and Physiological Psychology,* 1957, **50**, 195–198.

Clayton, K. N. Overlearning and reversal of a spatial discrimination by rats. *Perceptual and Motor Skills,* 1963, **17**, 83–85. (a)

Clayton, K. N. Reversal performance by rats following overlearning with and without irrelevant stimuli. *Journal of Experimental Psychology,* 1963, **66**, 255–259. (b)

D'Amato, M. R., & Jagoda, H. Effects of extinction trials on discrimination reversal. *Journal of Experimental Psychology,* 1960, **59**, 254–260.

D'Amato, M. R., & Jagoda, H. Analysis of the role of overlearning in discrimination reversal. *Journal of Experimental Psychology,* 1961, **61**, 45–50.

D'Amato, M. R., & Jagoda, H. Overlearning and position reversal. *Journal of Experimental Psychology,* 1962, **64**, 117–122.

D'Amato, M. R., & Schiff, D. Further studies of overlearning and position reversal learning. *Psychological Report,* 1964, **14**, 380–382.

D'Amato, M. R., & Schiff, D. Overlearning and brightness discrimination reversal. *Journal of Experimental Psychology,* 1965, **69**, 375–381.

Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review,* 1959, **66**, 183–201.

Eckstrand, G. A., & Wickens, D. D. Transfer of perceptual set. *Journal of Experimental Psychology,* 1954, **47**, 274–278.

Ehrenfreund, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *Journal of Comparative and Phsiological Psychology,* 1948, **41**, 408–422.

Erlebacher, A. Reversal learning in rats as a function of percentage of reinforcement and degree of learning. *Journal of Experimental Psychology,* 1963, **66**, 84–90.

Estes, W. K. Learning theory and the new "mental chemistry." *Psychological Review,* 1960, **67**, 207–223.

Estes, W. K. All-or-none processes in learning and retention. *American Psychologist,* 1964, **19**, 16–25.

Ferster, C. B. The effect on extinction responding of stimuli continuously present during conditioning. *Journal of Experimental Psychology,* 1951, **42**, 443–449.

Galanter, E., & Bush, R. R. Some T-maze experiments. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: University Press, 1959. Pp. 265–289.

Gatling, F. P. A study of the continuity of the learning process as measured by habit reversal in the rat. *Journal of Comparative and Physiological Psychology,* 1951, **44**, 78–83.

Goodrich, K. P., Ross, L. E., & Wagner, A. R. An examination of selected aspects of the continuity and noncontinuity positions in discrimination learning. *Psychological Records,* 1961, **11**, 105–117.

Goodwin, W. R., & Lawrence, D. H. The functional independence of two discrimination habits associated with a constant stimulus situation. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 437–443.

Goss, A. E. Verbal mediating responses and concept formation. *Psychological Review,* 1961, **68**, 248–274.

Grice, G. R. The aquisition of a visual discrimination habit following response to a single stimulus. *Journal of Experimental Psychology,* 1948, **38**, 633–642.

Grice, G. R. The acquisition of a visual discrimination habit following extinction of response to one stimulus. *Journal of Comparative and Physiological Psychology,* 1951, **44**, 149–153.

Guttman, N. Laws of behavior and facts of perception. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 5. New York: McGraw-Hill, 1963. Pp. 114–178.

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *Journal of Experimental Psychology,* 1956, **51**, 79–88.

Harlow, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 492–537.

Harrow, M. Stimulus aspects responsible for the rapid acquisition of reversal shifts in concept formation. *Journal of Experimental Psychology,* 1964, **67**, 330–334.

Harrow, M., & Friedman, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *Journal of Experimental Psychology,* 1958, **55**, 592–597.

Heinemann, E. G., & Rudolph, R. L. The effect of discriminative training on the gradient of stimulus generalization. *American Journal of Psychology,* 1963, **76**, 653–658.

Hicks, L. H. Effects of overtraining on acquisition and reversal of place and response learning. *Psychological Reports,* 1964, **15**, 459–462.

Hill, W. F., & Spear, N. E. Extinction in a runway as a function of acquisition level and reinforcement percentage. *Journal of Experimental Psychology,* 1963, **65**, 495–500. (a)

Hill, W. F., & Spear, N. E. A replication of overlearning and reversal in a T maze. *Journal of Experimental Psychology,* 1963, **65**, 317. (b)

Hill, W. F., Spear, N. E., & Clayton, K. N. T-maze reversal after several different overtraining procedures. *Journal of Experimental Psychology,* 1962, **64**, 533–540.

Hoffeld, D. R., Kendall, S. B., Thompson, R. F., & Brogden, W. J. Effect of amount of preconditioning training upon the magnitude of sensory preconditioning. *Journal of Experimental Psychology,* 1960, **59**, 198–204.

House, B. J., & Zeaman, D. Transfer of a discrimination from objects to patterns. *Journal of Experimental Psychology,* 1960, **59**, 298–302.

House, B. J., & Zeaman, D. Reversal and nonreversal shifts in discrimination learning of retardates. *Journal of Experimental Psychology,* 1962, **63**, 444–451.

Hughes, C. L., & North, A. J. Effect of introducing a partial correlation between a critical cue and a previously irrelevant cue. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 126–128.

Isaacs, I. D., & Duncan, C. P. Reversal and nonreversal shifts within and between dimensions in concept formation. *Journal of Experimental Psychology,* 1962, **64**, 580–585.

Ison, J. R. Experimental extinction as a function of number of reinforcements, *Journal of Experimental Psychology,* 1962, **64**, 314–317.

Ison, J. R., & Birch, D. T-maze reversal following differential end box placement. *Journal of Experimental Psychology,* 1961, **62**, 200–202.

Ison, J. R., & Cook, P. E. Extinction performance as a function of incentive magnitude and number of acquisition trials. *Psychonomic Science,* 1964, **1**, 245–246.

Iwahara, S. On overlearning and shift learning by white rats. *Procedings of the 23rd Japanese Psychological Association Meeting,* 1959.

James, W. *Principles of psychology.* New York: Holt, 1890.

Jeeves, M. A., & North, A. J. Irrelevant or partially correlated stimuli in discrimination learning. *Journal of Experimental Psychology,* 1956, **52**, 90–94.

Jenkins, H. M., & Harrison, R. H. Effect of discrimination training on auditory generalization. *Journal of Experimental Psychology,* 1960, **59**, 246–253.

Jenkins, H. M., & Harrison, R. H. Generalization gradients of inhibition following auditory discrimination learning. *Journal of the Experimental Analysis of Behavior,* 1962, **5**, 435–441.

Jensen, G. D., & Cotton, J. W. Running speed as a function of stimulus similarity and number of trials. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 474–476.

Johnson, E. E. The role of motivational strength in latent learning. *Journal of Comparative and Physiological Psychology,* 1952, **45**, 526–530.

Kelleher, R. T. Discrimination learning as a function of reversal and nonreversal shifts. *Journal of Experimental Psychology,* 1956, **51**, 379–384.

Kendler, H. H., & D'Amato, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation. *Journal of Experimental Psychology,* 1955, **49**, 165–174.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychological Review,* 1962, **69**, 1–16.

Kendler, T. S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology,* 1959, **58**, 55–60.

Kendler, T. S., Kendler, H. H., & Wells, D. Reversal and nonreversal shifts in nursery school children. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 83–88.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

Komaki, J. The facilitative effect of overlearning in discrimination learning by white rats. *Psychologia,* 1961, **4**, 28–35.

Komaki, J. Reversal retardation through introducing forced nonrewarding trials in overtraining by white rats. *Annual of Animal Psychology,* 1962, **12**, 1–10.

Krechevsky, I. A study of the continuity of the problem-solving process. *Psychological Review,* 1938, **45**, 107–133.

Krechevsky, I., & Honzik, C. H. Fixation in the rat. *University of California Publications in Psychology,* 1932, **6**, 13–26.

Lashley, K. S. *Brain mechanisms and intelligence.* Chicago: Univer. Chicago Press, 1929.

Lashley, K. S. The mechanism of vision: XV. Preliminary studies of the rat's capacity for detail vision. *Journal of General Psychology,* 1938, **18**, 123–193.

Lashley, K. S. An examination of the continuity theory as applied to discriminative learning. *Journal of General Psychology,* 1942, **26**, 241–265.

Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. *Psychological Review,* 1946, **53**, 72–87.

Law, O. I. Preference in the rat for vertical or horizontal stripes after training on a white-black discrimination. *American Journal of Psychology,* 1954, **67**, 714–716.

Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology,* 1949, **39**, 770–784.

Lawrence, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *Journal of Experimental Psychology,* 1950, **40**, 175–188.

Lawrence, D. H. The transfer of a discrimination along a continuum. *Journal of Comparative and Physiological Psychology,* 1952, **45**, 511–516.

Lawrence, D. H. The applicability of generalization gradients to the transfer of a discrimination along a continuum. *Journal of General Psychology,* 1955, **52**, 37–48.

Lawrence, D. H. The nature of a stimulus: Some relationships between learning and perception. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 5. New York: McGraw-Hill, 1963. Pp. 179–212.

Levine, M. A model of hypothesis behaviour in discrimination learning set. *Psychological Review,* 1959, **66**, 353–366.

Lovejoy, E. P. An analysis of the overlearning reversal effect. *Psychological Review,* 1965, **72**, in press.

MacCaslin, E. F., Wodinsky, J., & Bitterman, M. E. Stimulus-generalization as a function of prior training. *American Journal of Psychology,* 1952, **65**, 1–15.

Mackintosh, N. J. The effects of overtraining on a reversal and a nonreversal shift. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 555–559.

Mackintosh, N. J. Direct transfer from a horizontal-vertical discrimination to a brightness discrimination in the rat. *Quarterly Journal of Experimental Psychology,* 1963, **15**, 212–213. (a)

Mackintosh, N. J. The effect of irrelevant cues on reversal learning in the rat. *British Journal of Psychology,* 1963, **54**, 127–134. (b)

Mackintosh, N. J. Extinction of a discrimination habit as a function of overtraining. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 842–847. (c)

Mackintosh, N. J. Overtraining and transfer within and between dimensions in the rat. *Quarterly Journal of Experimental Psychology,* 1964, **16**, 250–256.

Mackintosh, N. J. The effect of attention on the slope of generalization gradients. *British Journal of Psychology,* 1965, **56**, 87–93. (a)

Mackintosh, N. J. Overtraining, extinction, and reversal in rats and chicks. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 31–36. (b)

Mackintosh, N. J. Overtraining, transfer to proprioceptive control, and position reversal. *Quarterly Journal of Experimental Psychology,* 1965, **17**, 26–36. (c)

Mackintosh, N. J. Transposition after "single stimulus" pretraining. *American Journal of Psychology,* 1965, **78**, 116–119. (d)

Mackintosh, N. J. Incidental cue learning in rats. *Quarterly Journal of Experimental Psychology,* in press.

Mackintosh, N. J., & Mackintosh, J. Reversal learning in *Octopus vulgaris Lamarck* with and without irrelevant cues. *Quarterly Journal of Experimental Psychology,* 1963, **15**, 236–242.

Mackintosh, N. J., & Mackintosh, J. The effect of overtraining on a nonreversal shift in *Octopus. Journal of Genetic Psychology,* 1964, **106**, 373–377.

McCain, G., & Garrett, B. L. Generalization to stimuli of different brightness in three straight-alley studies. *Psychological Reports,* 1964, **15**, 368–370.

McCullogh, T. L., & Pratt, J. G. A study of the presolution period in weight discrimination by white rats. *Journal of Comparative and Physiological Psychology,* 1934, **18**, 271–290.

McGuigan, F. J., & Crockett, F. Evidence that the secondary reinforcing stimulus must be discriminated. *Journal of Experimental Psychology,* 1958, **55**, 184–187.

Madison, H. L. Experimental extinction as a function of number of reinforcements. *Psychological Reports,* 1964, **14**, 647–650.

Mahut, H. The effect of stimulus position on visual discrimination by the rat. *Canadian Journal of Psychology,* 1954, **8**, 130–138.

Miller, N. E., & Dollard, J. C. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

North, A. J. Acquired distinctiveness of form stimuli. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 339–341.

North, A. J., & Clayton, K. N. Irrelevant stimuli and degree of learning in discrimination learning and reversal. *Psychological Reports,* 1959, **5**, 405–408.

North, A. J., & Stimmel, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychological Reports,* 1960, **6**, 227–234.

Notterman, J. M. A study of some relations among aperiodic reinforcement, discrimination training, and secondary reinforcement. *Journal of Experimental Psychology,* 1951, **41**, 161–169.

Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

Paul, C., & Kesner, R. Effects of overlearning trials upon habit reversal under conditions of aversive stimulation. *Psychological Record,* 1963, **13**, 361–363.

Pavlov, I. P. *Conditioned reflexes.* Oxford: Univer. Press, 1927.

Perkins, C. C., Hershberger, W. A., & Weyant, R. G. Difficulty of a discrimination as a determiner of subsequent generalization along another dimension. *Journal of Experimental Psychology,* 1959, **57**, 181–186.

Prokasy, W. F., & Hall, J. F. Primary stimulus generalization. *Psychological Review,* 1963, **70**, 310–322.

Pubols, B. H., Jr. The facilitation of visual and spatial discrimination reversal by overlearning.

Journal of Comparative and Physiological Psychology, 1956, **49**, 243–248.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *Journal of Experimental Psychology*, 1953, **46**, 107–112.

Restle, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 329–343.

Reynolds, G. S. Attention in the pigeon. *Journal of the Experimental Analysis of Behavior*, 1961, **4**, 203–208.

Ritchie, B. F., Ebeling, E., & Roth, W. Evidence for continuity in the discrimination of vertical and horizontal patterns. *Journal of Comparative and Physiological Psychology*, 1950, **43**, 168–180.

Ross, L. E. The effect of equal reinforcement of the positive and negative discriminanda of a learned discrimination. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 260–266.

Ross, L. E. The response to previous discriminanda during the learning of a new problem. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 944–946.

Sasaki, M. The effect of overtraining upon transfer in discrimination learning. *Proceedings of the 24th Japanese Psychological Association Meeting*, 1960, 199–200.

Shepp, B. E., & Eimas, P. D. Intradimensional and extradimensional shifts in the rat. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 357–361.

Siegel, P. S. The role of absolute initial response strength in simple trial-and-error learning. *Journal of Experimental Psychology*, 1945, **35**, 199–205.

Siegel, S., & Wagner, A. R. Extended acquisition training and resistance to extinction. *Journal of Experimental Psychology*, 1963, **66**, 308–310.

Spence, K. W. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427–449.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, 1937, **44**, 430–444.

Spence, K. W. Continuous versus non-continuous interpretations of discrimination learning. *Psychological Review*, 1940, **47**, 271–288.

Spence, K. W. An experimental test of continuity and non-continuity theories of discrimination learning. *Journal of Experimental Psychology*, 1945, **35**, 253–266.

Spence, K. W. *Behavior theory and learning*. Englewood Cliffs: Prentice-Hall, 1960.

Sperling, Sally E. Reversal learning and resistance to extinction: A review of the rat literature. *Psychological Bulletin*, 1965, **63**, 281–297.

Stevenson, H. W., Iscoe, I., & McConnell, C. A developmental study of transposition. *Journal of Experimental Psychology*, 1955, **49**, 278–280.

Sutherland, N. S. Stimulus analysing mechanisms. In, *Proceedings of a symposium on the mechanization of thought processes*. Vol. 2. London: Her Majesty's Stationery Office, 1959. Pp. 575–609.

Sutherland, N. S. The learning of discriminations by animals. *Endeavour*, 1964, **23**, 148–152.

Sutherland, N. S., & Mackintosh, J. Discrimination learning: Non-additivity of cues. *Nature*, 1964, **201**, 528–530.

Sutherland, N. S., Mackintosh, N. J., & Mackintosh, J. Simultaneous discrimination training of *Octopus* and transfer of discrimination along a continuum. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 150–156.

Sutherland, N. S., Mackintosh, N. J., & Mackintosh, J. Shape and size discrimination in *Octopus:* The effects of pretraining along different dimensions. *Journal of Genetic Psychology*, 1965, **107**, 1–10.

Sutherland, N. S., Mackintosh, N. J., & Wolfe, J. B. Extinction as a function of the order of partial and consistent reinforcement. *Journal of Experimental Psychology*, 1965, **69**, 56–59.

Takemura, K. The effect of overlearning on the subsequent task with different type of solution. In, *Proceedings of the 25th Japanese Psychological Association Meeting*, 1961, 132.

Terrace, H. S. Errorless transfer of discrimination across two continua. *Journal of Experimental Analysis of Behavior*, 1963, **6**, 223–232.

Theios, J., & Blosser, D. The overlearning reversal effect and magnitude of reward. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 252–257.

Thompson, M. E. Stimulus generalization of an instrumental response learned under distributed practice. *Psychological Reports*, 1962, **11**, 471–476.

Uhl, C. N. Effects of overtraining on reversal and nonreversal discrimination shifts in a free operant situation. *Perceptual and Motor Skills*, 1964, **19**, 927–934.

Wagner, A. R. Overtraining and frustration. *Psychological Reports*, 1963, **13**, 717–718.

Walk, R. D. Effect of discrimination reversal on human discrimination learning. *Journal of Experimental Psychology*, 1952, **44**, 410–419.

Wortz, E. C., & Bitterman, M. E. On the effect of an irrelevant relation. *American Journal of Psychology*, 1953, **66**, 491–493.

Wyckoff, L. B., Jr. The role of observing responses in discrimination learning. Part I. *Psychological Review*, 1952, **59**, 431–442.

Youniss, J., & Furth, H. G. Reversal learning in children as a function of overtraining and delayed transfer. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 155–157. (a)

Youniss, J., & Furth, H. G. Reversal performance of children as a function of overtraining and response conditions. *Journal of Experimental Child Psychology*, 1964, **1**, 182–188. (b)

Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook in mental deficiency: Psychological theory and research*. New York: McGraw-Hill, 1963. Pp. 159–223.

(Received February 23, 1965)

Psychological Bulletin
1965, Vol. 64, No. 2, 151–152

A NOTE ON REPEATED MEASURES IN THE STUDY OF SHORT-TERM MEMORY

LLOYD R. PETERSON[1]

Indiana University

Keppel's (1965) objection to averaging over repeated measures is discussed. Several possibilities for minimizing proactive interference are suggested, including increasing time between tests. Consideration is given to acceptance of a constant amount of proactive interference over conditions in which other variables are studied. The advantages of using repeated measures in regard to efficiency as well as interpretation are noted.

Keppel (1965) has written a stimulating discussion of some methodological problems in the study of short-term verbal retention. One problem concerns the use of repeated measures, and his conclusions in this regard may be taken as a virtual proscription on averaging over repeated tests on the same subject. He says, ". . . any complete study of STM must involve a manipulation of the amount of prior-learned material [p. 6]." In the discussion that follows it becomes clear that this means testing large numbers of subjects "to allow independent-group comparisons on the first retention test and a systematic introduction of subsequent tests under the same and different conditions."

The implications of this position for research in the area are serious. The value of a large number of studies already reported becomes questionable, and the future is bleak, since the large numbers of subjects required for experiments of the type that Keppel endorses will make research slow and expensive. Further, the suggested solution will render difficult any conclusions about the nature of retention functions for individual subjects, for averaging between subjects takes the place of averaging within subjects. An instance of the importance of individual differences in short-term memory is seen in a report by Pollack (1963). In an analysis of repeated measurements Pollack compared the highest third of his subjects (in terms of total performance) with the middle and lowest third. Although the differences in the subjects' capacities to immediately recall six digits were small, the differences when one extra digit intervened between presentation and recall were very large. The high scoring group dropped from 98% correct without the extra digit to 95% with the extra digit. On the other hand, the low scoring subjects dropped from 83% to 35% correct with the addition of the digit. The curves obtained in a parametric study of number of interfering digits and duration of additional delay were plotted separately for high, middle, and low subjects, and the differences in shapes of the curves were marked. Such differences would have been overlooked without averages over repeated measurements of the same subjects.

It should be noted that Keppel's (1965) concern over repeated measures stems from studies (Keppel & Underwood, 1962) in which a progressive decrement has been observed over the first few tests with a technique in which recall of a trigram is measured after an interval filled with irrelevant verbal behavior. This is an important finding which must be considered in accounting for short-term forgetting. However, when this decrement is not the object of study for its own sake, there are ways of minimizing it to permit study of other variables. To the extent that this decrement is due to proactive interference, the effect should be maximized for retention intervals of long duration, since proactive interference increases with the retention interval (Keppel & Underwood, 1962). Conversely, proactive interference is minimized in that class of studies in which recall is tested shortly after presentation, as in studies of immediate memory span. In the testing of longer retention intervals, proactive interference can be reduced by lengthening the interval between tests. In an extreme case, Loess (1964) found that after a period of three weeks no effect of a previous session remained. In dealing with more workable intervals between tests, Peterson and Gentile (1965) have found a significant reduction in the effect of previous tests when the interval from the signal for recall in one test to the beginning of the next test was lengthened from 5 to 16 seconds, and a further reduction from 16 to 91 seconds. In the Keppel and Underwood studies (1962) the intertest interval was not specified, but Keppel[2] indicated that it was 10 seconds. This is shorter than the 15 seconds of the initial studies using this technique (Peterson & Peterson, 1959). The marked progressive decrement as a function of previous tests, found by Keppel and Underwood in their third experiment (1962), may well have been influenced by this short intertest interval as well as by the bias that Keppel mentioned. In other experiments, as Keppel noted, a steady state has quickly been reached.

---

[1] This note was written at the Institute of Human Learning, University of California, Berkeley.

Since a steady state is reached after a few initial tests, the experimenter can adopt the strategy of accepting a small constant amount of proactive interference, and allow performance to stabilize during initial practice tests. The experimenter will recognize that his results may not generalize without modification to the case where proactive interference is of less effect. This was done in an experiment yielding a curve which Keppel says should be labeled the "RI [retroactive inhibition] work curve." In so labeling a curve from a study by Peterson, Saltzman, Hillner, and Land (1962), Keppel implicitly acknowledges the legitimacy of the principle of balancing out a constant amount of an unwanted variable. His RI work curve was obtained under conditions where varying numbers of previous pairs were given before the presentation of a tested pair. The first pair in a block of presentations was a dummy used to avoid the marked decrement commonly found between the first and second pairs in a sequence. Other differences in numbers of previous pairs were balanced out by the experimental design. The curve could be called an RI work curve under conditions of a constant amount of proactive interference. The situation involves a kind of mixing of proactive and retroactive interference similar to that obtaining within a list of pairs being learned over trials. The chief difference is that retention intervals for individual pairs were not averaged as in list learning, but each length of interval was treated separately.

REFERENCES

Keppel, G. Problems of method in the study of short-term memory. *Psychological Bulletin*, 1965, **63**, 1–13.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Loess, H. Proactive inhibition in short-term retention. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 362–368.

Peterson, L. R., & Gentile, A. Proactive interference as a function of time between tests. *Journal of Experimental Psychology*, 1965, **70**, in press.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

Peterson, L. R., Saltzman, D., Hillner, K., & Land, V. Recency and frequency in paired-associate learning. *Journal of Experimental Psychology*, 1962, **63**, 396–403.

Pollack, I. Interference, rehearsal, and short-term retention of digits. *Canadian Journal of Psychology*, 1963, **17**, 380–392.

(Received March 26, 1965)

---

[2] Personal communication, 1965.

VOL. 64, NO. 3 SEPTEMBER 1965

Psychological Bulletin

EQUIDISTANCE TENDENCY AND ITS CONSEQUENCES

WALTER C. GOGEL[1]

Civil Aeromedical Research Institute, Oklahoma City

The equidistance tendency is the tendency for objects or other inhomogeneities in the field of view to appear at the some distance as each other with the strength of this tendency being inversely related to directional separation. The evidence for the existence of the equidistance tendency and for its ability to modify the perceived depth resulting from size or stereoscopic cues is reviewed. The equidistance tendency is discussed as a disturbing factor in visual experimentation and as a necessary factor in the understanding of Emmert's law, the moon illusion, and similar phenomena. Several possible explanations for the equidistance tendency are evaluated briefly in terms of the range of phenomena with which it is identified.

The perception of volume seems to be the usual perception resulting from homogeneous visual stimulation (Natsoulas, 1963). It follows that with homogeneous visual fields the presence of specific cues is not a necessary condition for the perception of depth. However, the same conclusion does not necessarily apply to the perception of depth between inhomogeneities in the visual field when these occur without the introduction of depth cues. Specifically, the question might be asked as to how two objects appear in depth with respect to each other when all cues which would indicate that they are at different distances are removed. It might seem reasonable to expect that, in the absence of depth cues, the perceived depth between two objects would be indeterminate or controlled by momentary attitudinal biases. According to this point of view, depth perception in the absence of visual cues would be erratic because there is no process which would result in one perception being more likely to occur than another. However, in spite of its apparent inevitability, this point of view is not the only one possible. Instead, it might be supposed that in the absence of visual cues the perceived depth between two objects would tend toward some specific perception which would be determined by organizing factors within the observer. It is the purpose of this paper to support this latter point of view by considering the evidence for a tendency to perceive objects as equidistant when effective cues as to their perceived depth are absent. Furthermore, it has been demonstrated that this tendency to perceptual equidistance can operate with some degree of effectiveness in situations in which not all depth cues are eliminated. As will be discussed, as a result of this characteristic, the equidistance tendency exercises an often unrecognized influence upon perceptions occurring under a variety of visual conditions.

EVIDENCE FOR THE EQUIDISTANCE TENDENCY

The equidistance tendency is the tendency for objects or parts of objects, in the absence of effective distance cues, to appear visually at the same distance as each other with the strength of this tendency being inversely related to the directional separation of the objects or parts (cf. Gogel, 1963). When two objects (a luminous square and a rectangle) physically located at different distances from the observer were viewed in the absence of distance cues, it was found that the two objects appeared to be at approximately the same distance (Gogel, 1956b, Experiment I). It was therefore hypothesized that in the absence of contrary visual information there was a tendency for objects to appear equidistant. The word "tendency" was used since, in general, it was found that considerable individual variation could occur in this response.

---

[1] Now at the University of California, Santa Barbara.

If it is assumed that the equidistance tendency is present between all objects in the field of view, it follows that, as the number of objects increases, the number of tendencies toward equidistance increases with some of these possibly being in conflict. For example, if two objects appear at different distances, a third object with no depth cues with respect to either of these objects cannot appear simultaneously to be at the distance of both objects. It was found under these conditions (Gogel, 1956b, Experiment II) that the third object appeared nearest in depth to that object to which it was directionally most adjacent. It was as though each of the two objects contributed an equidistance effect with respect to the third object, but the object directionally most adjacent to the third object contributed most. Thus, a tendency to perceptual equidistance occurs between a pair of objects or contours with the strength of this tendency being inversely related to directional visual separation. These conclusions are further substantiated by an experiment (Gogel, 1956b, Experiment III) which investigated the perceptual effect of a row of luminous cylinders receding in depth from near right to far left (or vice versa) upon the perceived distance of a single luminous cylinder. Again, in each of the situations used, the single cylinder was on the average closer in apparent depth to the directionally adjacent rather than to the directionally displaced cylinders in the row of cylinders. This occurred in spite of the tendency for the relative size cue between cylinders to localize the single cylinder midway in depth in the row.

In the experiments discussed above, the object whose apparent depth position was modified by the equidistance tendency was viewed monocularly with the remaining object or objects viewed binocularly. This arrangement was used because it provided certain conveniences of measurements and experimental manipulation. However, it is not intended that these results should apply only to situations involving a combination of monocularly and binocularly viewed objects (cf. Gogel, 1956a, 1956b, 1963).

Roelofs and Zeeman (1957) investigated, by means of stereograms, the perceived size and distance of monocularly viewed objects presented simultaneously with binocularly viewed objects. Only the results concerning perceived distance are pertinent for the present study. In one series of stereograms, two monocularly observed and two binocularly observed circles were viewed with one of the binocularly observed circles being stereoscopically more distant than the other. One of the monocularly seen circles was closer on the stereogram to the more distant of the binocularly seen circles and the other closer to the less distant of the binocularly seen circles. The results were that the monocularly seen circle which was closer on the stereogram (directionally closer in the visual field) to the stereoscopically farther, binocularly seen circle usually was perceived to be the more distant of the two monocularly seen circles. It was concluded that the objects seen binocularly influenced the apparent distance of the objects seen monocularly, with the strength of this influence being greater with less separation between them in the visual field. This conclusion is clearly in agreement with the demonstrations of the equidistance tendency cited above.

In a supplementary series of stereograms, Roelofs and Zeeman presented a binocularly viewed and a monocularly viewed circle. The monocularly viewed circle tended to be perceived as slightly nearer than the binocularly viewed circle. This difference, although minute and susceptible to suggestion, (Roelofs & Zeeman, 1957, p. 199) also appeared in a later study (Roelofs, 1961). Additional tendencies were explored in both series of stereograms and in the later study. There seems to be a tendency, other things being equal, to see objects in the upper, right, and temporal half of the visual field as more distant than objects in the opposite positions. Roelofs (1961), while asserting the validity of the equidistance tendency, pointed out from these and other demonstrations that some of the positional depth tendencies which occur cannot be subsumed under the equidistance tendency. Perhaps it is reasonable to expect that the equidistance tendency is not the only such tendency in the visual field. It is the point of view of this paper, however, that the equidistance tendency, since it applies throughout the visual field, represents a general perceptual factor upon which other tendencies must be superimposed.

The equidistance tendency has been demonstrated to apply when all the objects involved are viewed monocularly. For this demonstration, two playing cards were used which appeared at different distances as a consequence of a difference in retinal size (Gogel, 1956a, pp. 4–6). A small disc was placed directionally closer to either the apparently nearer or the apparently farther card or midway between the cards in direction. The average results indicated that the disc appeared closer in depth to the card to which it was directionally closer, and it appeared approximately midway between the cards in depth when it was midway between them in direction. In a further experiment (Gogel, 1956a, pp. 7–14), the apparent path of movement of a monocularly observed disc attached to a monocularly observed rotating trapezoidal window was analyzed, using the generalization that the resultant strength of the equidistance tendency with respect to the various parts of the window would vary as a function of the changing directional orientation of the disc with respect to the window. This latter experiment not only demonstrates that binocular observation of some of the objects is not a necessary condition in order for the equidistance tendency to be effective, it also supports the concept that the equidistance tendency applies to parts of objects as well as to objects per se. All objects, contours, and all parts of objects in the visual field both exert, and are subject to, a visual equidistance effect with respect to all other objects, contours, and parts of objects, with the strength of the equidistance tendency between any two visual inhomogeneities being inversely related to their difference in visual direction. It follows, for example, that all parts of a figure slanted in depth will tend to appear in an apparent frontoparallel plane in the absence of strong cues to slant.

MODIFICATIONS BY THE EQUIDISTANCE TENDENCY OF JUDGMENTS INVOLVING DEPTH CUES

Modifications of Judgments Involving Stereoscopic Cues

The equidistance tendency sometimes can modify depth perceptions even in situations in which stereoscopic cues to the apparent depth of the objects are available. This has been demonstrated in a study by Gogel, Brune, and Inaba (1954) in which the conditions were such that the strength of the equidistance tendency was near a maximum relative to that of the stereoscopic cue. That the equidistance tendency can modify a stereoscopic judgment under other viewing conditions is suggested by an experiment by Harker (1962) in which a grid was stereoscopically oriented at various slants. The slant of the grid would be expected to induce cyclorotation of the eyes resulting in a reduction in the apparent tilt of the grid. The equidistance tendency also would be expected to act in the direction of reducing the apparent tilt by tending to make all parts of the grid appear at the same distance. Harker found that only some of the reduction in perceived tilt could be attributed to cyclorotation with the remainder attributed to the action of the equidistance tendency. Simon (1936), in a series of experiments, found that a binocularly observed outline circle physically slanted in depth would change its apparent orientation under reduced conditions of observation. Simon suggested that the perceived orientation (and perceived distance) of such objects would be along an ellipsoid with the observer at a focal point. Clearly the observations upon which this conclusion was based are very similar to those expected from the equidistance tendency and probably represent a modification by the equidistance tendency of judgments involving stereoscopic cues.

Several studies concerned with aftereffects also may support (at least inferentially) the notion that the equidistance tendency can modify stereoscopic judgments. In studies by Köhler and Emory (1947), Bergman and Gibson (1959), and Howard and Templeton (1964), it was found that the prior presentation of inspection figures physically arranged in depth and viewed binocularly affected the perceived depth involved in subsequently presented test figures. Köhler and Emory related their results to satiation theory. Bergman and Gibson described their results as a negative aftereffect. A physically slanted, textured surface with either steady or moving fixation appeared less slanted with continued observation. A subsequently presented surface in the same locality had an error in perceived slant which was opposite to that found with the original presentation. Howard and Templeton explained their results by assuming that the inspection stimulus appeared less three dimensional than it actually was (the equidistance tendency) and, consequently, that the subsequent stimuli appeared distorted in depth in a direction opposite to that of the distortion in the inspection display. Since at least some of the effects found in these studies are attributable to the equidistance tendency, it is not only likely that the equidistance tendency, under certain conditions, can modify stereoscopic judgments, but also that equidistance effects occurring in one presentation can extend to other presentations.

Modifications of Judgments Involving Size Cues

It was noted that the equidistance tendency could be effective even under circumstances in which it opposed the size cue to relative distance (Gogel, 1956b, Experiment III). An experiment by Gogel and Harker (1955) supports the same conclusion. It was found that the perceived depth between two playing cards of different retinal size increased with an increase in their lateral separation. This increase in perceived depth could be attributed either to an increase in the strength of the size cue with increased directional separation or to a decrease in the strength of the equidistance tendency with increased directional separation. In a later experiment (Gogel, 1956b, Experiment IV) using the larger lateral separation, the equidistance tendency resulting from the introduction of an additional object opposed the perceived depth normally occurring between the cards. It was found that the additional object clearly modified the perceived depth between the cards. It appears likely, therefore, that the increase in apparent depth with increased lateral separation which occurred in the previous experiment represented a decrease in the strength of the equidistance tendency, not an increase in the strength of the size cues. It seems that the equidistance tendency can modify a perceived depth even though opposed by the size cues which occur from familiar or nonfamiliar objects.

In the study by Bergman and Gibson (1959) the inspection figure consisted of a textured surface (burlap) which was viewed through an aperture in order to eliminate any perception of the edges of the surface. The texture of the slanted surface provided a gradient of size changes (size cues) for the perception of slant. There was a tendency for this surface when viewed either monocularly or binocularly to appear frontal with continued observation. From the viewpoint of the present paper, this result represents a modification by the equidistance tendency of the depth perception resulting from size cues alone or from size and binocular disparity cues together.

OTHER INSTANCES OF THE EFFECTS OF THE EQUIDISTANCE TENDENCY

Visual conditions in which the equidistance tendency would have some perceptual effect occur frequently. Therefore, it is to be expected that phenomena which can be attributed to this tendency would have been noted either explicitly or implicitly in a variety of situations. This is indeed the case, as the following instances demonstrate. Judd has reported (1898) that white threads presented against a black background and located at different distances from an observer appeared to be equidistant when viewed monocularly, whereas with binocular observation they were clearly perceived to be at different distances. In a discussion of the perceived depth between diplopic images and a binocularly fixated object, Judd stated: "And we may assert it as a general principle that in monocular vision objects tend toward a single plane, the plane determined by binocular factors [p. 293]." Katz (1935, pp. 71–74) noted that the apparent location and orientation of film color is at least partially determined by the location and orientation of the aperture through which it is viewed. Simon (1936) discussed some of the early observations on perceived distortions which related to his own experiments and, it would seem, also relate to the equidistance tendency. Gibson and Carel (1952) reported that a display of stationary lights in a reduced cue situation was perceived as occupying a frontoparallel plane. Also, the results from a number of experiments concerning the perception of slant (Beck & Gibson, 1955; Bergman & Gibson, 1959; Clark, 1953; Clark, Smith, & Rabe, 1955, 1956a, 1956b; Gibson, 1950b; Gibson & Gibson, 1957; Gruber & Clark, 1956; Smith, 1956, 1959, 1964a, 1964b, 1964c; Stravrianos, 1945) or curvature (Smith & Smith, 1957, 1961) indicate that, under somewhat reduced cue situations, the slant or curvature of a stationary surface is usually underestimated, that is, the perceived slant or perceived curvature of the surface tends toward a frontoparallel plane. It will be noted that these results represent phenomena in agreement with the equidistance tendency in that they involve the tendency for all parts of the surface to appear equidistant.

Epstein and Park (1964), in considering studies involving the effect of the depth orientation of the ground upon the apparent depth orientation of the figure, concluded that "There is considerable evidence that the slant of a figure tends to be assimilated to the slant of its ground [p. 184]." From the point of view of the present paper, the equidistance tendency is responsible for this process. Under conditions such that cues of the orientation of the figure are totally reduced, the figure will appear on the ground according to the equidistance tendency; that is, each part of the figure will appear at the depth of its directionally adjacent portion of the ground. If the reduction in cues is only partial, the equidistance tendency will be able to only partially displace the apparent slant of the figure in the direction of the apparent slant of the ground. If cues as to the orientation of both the figure and ground are reduced, both the figure and ground will tend to appear in a frontoparallel plane.

From the previous discussion, it is clear that the possible effect of the equidistance tendency must be evaluated in many experimental situations involving either a complete or a partial reduction in depth cues. If the effect of the equidistance tendency is not adequately considered in such situations, the interpretation concerning the events determining the perceptions can be seriously in error. Also, however, as has been noted, the equidistance tendency can be useful, as well as being a source of distortion of experimental data. Harker (1962) and Howard and Templeton (1964) have applied the equidistance tendency toward an explanation of binocular depth contrast (Werner, 1937, 1938), and Gogel (1964a) used the equidistance tendency in order to modify perceived position without varying binocular stimulation in a study concerned with the perception of depth from binocular disparity. Another possible application of the equidistance tendency is as an explanation of the ability of observers to make relative retinal size judgments under reduced cue conditions with monocular observation and the successive or simultaneous presentation of stimuli (cf. Epstein, Park, & Casey, 1961; Hartman, 1964). Contrary to the discussion of this problem by Epstein, Park, and Casey (1961, p. 493) the equidistance tendency has been demonstrated to occur between simultaneously presented objects when all, rather than only some of these objects, are monocularly observed (Gogel, 1956a). Furthermore, there is evidence that an equidistance effect can occur between directionally adjacent but successively presented stimuli. This is indicated by a study by Gruber, King, and Link (1963) in which a moon illusion was demonstrated using two artificial moons and the prior presentation of a luminous ceiling. It can be inferred from the ratio of the perceived sizes of the moons that the artificial moon which was directionally adjacent to a more distant portion of the ceiling, in general, appeared to be the more distant moon. However, even under reduced conditions of observation, the equidistance tendency may be ineffective when large directional separations occur between the successively presented objects (cf. Rock & McDermott, 1964).

APPLICATIONS OF THE EQUIDISTANCE TENDENCY

Application to the Cue of Relative Direction

A monocularly observed object which is viewed against a surface extending in depth will be perceived as resting on the surface. If the more distant portions of the surface are higher in the field of view, moving the object upward will make the object appear to be more distant. This was illustrated in a figure by Gibson (1950a, Fig. 72) in which two rectangles of different sizes were made to appear reversed in size by being placed at different heights and hence appeared at different distances in the visual scene. The same phenomenon has recently been used by McDonald and O'Hara (1964) in an investigation of the relation between perceived size and distance. The basis of this phenomenon is the visual interruption of a portion of the surface by the interposition of the bottom edge of the object between this portion of the surface and the eye of the observer. Gibson (1950a, p. 178) described this factor as the optical contact of the object with its background. The question to be answered is why the interposition results in a rectangle, for example, being perceived as resting on that portion of the surface with which its bottom edge has a common visual direction. A possible answer is that this phenomenon is an example of the equidistance tendency. An equidistance tendency is expected to occur between the rectangle and all parts of the surface. However, the strength of this tendency will be greatest with respect to the portion of the surface having the same visual direction as the rectangle. Therefore, under the conditions that cues to the actual depth position of the rectangle are restricted or eliminated, the equidistance tendency should perceptually localize the rectangle at or near the position where the line of sight from the eyes to the bottom of the rectangle intersects the surface.

It is of interest to note that if the line of sight to the bottom of a monocularly observed rectangle were higher than the most distant portion of a floor surface, the rectangle would appear (according to the equidistance tendency) to be at the more distant part of the scene, since the rectangle is closest to this part in direction (cf. Gibson, 1950a, Fig. 70). This observation illustrates the advantage of classifying the phenomenon with the equidistance tendency rather than with the factor of optical contact. If the surface extending in depth were above rather than below eye level (a ceiling rather than a floor), an upward movement of the rectangle would be perceived as resulting in a decrease rather than an increase in distance. Similarly, if the surface were a left or a right wall, a lateral movement of the rectangle would result in a change in the perceived distance. Neither height in the visual field nor optical contact with the surface adequately explains the range of possible phenomena. All these phenomena, however, can be considered as part of the more general perceptual event described as the equidistance tendency.

According to evidence presented in this paper, in the absence of other definitive visual cues an object will tend to be perceptually localized in a visual field by the equidistance tendency. The final equidistance effect in this case is the resultant of all the equidistance tendencies between the particular object and all contours and other objects in the visual field, with the most adjacent contours or objects having the greatest influence. Such an effect is clearly of consequence when using a comparison field to measure the perceived size or distance of a standard (experimental) object. Experimental results which are due to the happenstance that the standard object has a particular direction (and with respect to the comparison field (and therefore is subject to a particular equidistance resultant) might be misinterpreted by the experimenter as having some other significance.

Application to Emmert's Law

As usually expressed, Emmert's law states that the perceived size of an afterimage is directly proportional to the distance of the surface on which it is perceived. Two questions are basic to the investigation of Emmert's law: (*a*) What factors determine the position of the afterimage in the visual field? (*b*) How does the position of the afterimage in the visual field determine its apparent size? The first of these two questions directly involves the equidistance tendency. A convenient method of demonstrating Emmert's law is to measure (by means of a comparison object) the apparent size of a foveal afterimage when the observer fixates one and then another of several surfaces at different distances. One question which might be raised regarding this procedure is why the afterimage appears at the distance of the fixated surface. This question is the same question as asking why the image from a movie projector (or any other projector) appears at the distance of the screen. The answer is found in the equidistance tendency. The afterimage appears at the distance of the object to which it is directionally most adjacent and the surface on which it is made to appear represents maximum directional adjacency, that is, directional identity. However, directional identity should not be an essential condition for the demonstration of Emmert's law. Consider the situation in which a vertical surface is presented at one of several distances from the observer. Suppose that the observer looks in a direction slightly above the surface so that the foveal afterimage appears slightly above this surface in direction. If the surface is directionally the most adjacent object to the afterimage, the afterimage, according to the equidistance tendency, will appear approximately at the distance of the surface. If the surface is successively presented at different distances while maintaining its directional adjacency with respect to the afterimage, the apparent size of the afterimage should change, just as it does when it appears upon the surface. By itself, the afterimage has no cues to relate it to the visual field. It appears at one or another distance in the visual field completely as a consequence of the equidistance tendency. It follows that the equidistance tendency is an essential part of any explanation of Emmert's law.

The question concerning the apparent size of the afterimage has at least two possible answers. It might be suggested that the perceived size of an afterimage is directly proportional to its perceived distance from the observer (Price, 1961). According to this suggestion, Emmert's law is a special case of the size-distance invariance hypothesis (cf. Epstein, Park, & Casey, 1961) and thus would be subject to the evidence for and against this hypothesis. The role of the equidistance tendency in this process would be to determine the apparent distance by perceptually localizing the afterimage with respect to other objects. Another possible answer can be considered by returning to the example of an afterimage appearing at the distance of a directionally adjacent (or directionally identical) vertical surface. It has been asserted (Gogel, 1964b) that the perceived size of an object subtending a particular visual angle is determined by the perceived size (S′) per unit of visual angle (θ) of objects in its apparent depth vicinity. The contribution of the equidistance tendency to this process would be to make the afterimage appear at the depth vicinity of the surface and thus determine that the S'/θ value of the surface would become that of the afterimage. Clearly, for either of these two possible explanations of the apparent size of afterimages, the localization of the afterimage in the visual field by means of the equidistance tendency is a necessary event.

Applications to the Moon Illusion

Conclusions which apply to the perceived size and perceived distance of afterimages should also apply to the perceived size and perceived distance of the moon, and hence to the moon illusion (cf. Kaufman & Rock, 1962b; King & Gruber, 1962). The moon has few cues to localize it in distance with respect to other objects. Therefore, as a consequence of the equidistance tendency, it should tend to appear at the distance of objects directionally adjacent to it. Upon appearing at a particular location, the moon might appear to be a particular size as a consequence of (*a*) its perceived distance (the size-distance invariance type of explanation), or (*b*) its relative retinal size as compared with familiar objects in its perceived environment (the S'/θ type of explanation). Clearly, with either type of explanation, the equidistance tendency is a necessary explanatory factor. It follows that the perceived size of the moon should be capable of being systematically changed by using the equidistance tendency to make it appear at different distances or at the distances of objects with different values of S'/θ. For example, suppose that an optically generated disc of light (a moon) with a diameter subtending a visual angle of .5 degree is presented directionally above a 3.5 inch high playing card located at a distance so that its height subtends 1.75 degrees of visual angle. No other objects except the playing card and moon are visible, and all cues to the depth between the moon and card are eliminated. Under these conditions, the equidistance tendency would result in the moon appearing at the distance of the playing card and according to the S'/θ type of explanation the moon would have a perceived diameter of about one inch. Thus, the moon can be made to appear any size by having it appear at the distance of objects having particular values of S'/θ (or at particular, perceived distances).[2] It can be concluded, therefore, that the observation that the horizon moon looks larger than the zenith moon is a happenstance of the particular visual conditions usually encountered, since the appropriate manipulation of the visual conditions should be able to destroy the illusion or reverse its direction.

The above explanation of the moon illusion is consistent with the results of the recent research on the moon illusion by Kaufman and Rock (1962a, 1962b), King and Gruber (1962), and Rock and Kaufman (1962) in which evidence is presented against the "angle-of-regard hypothesis" of the moon illusion summarized by Boring (1943). However, unlike the S'/θ aspect of the explanation asserted in the present paper, Rock and Kaufman (1962) concluded that, in general, the perceived size of the moon is determined by its perceived distance and not by its comparison with surrounding objects. The latter explanation is rejected because terrain modifications can affect the perceived size of the moon and also because the moon illusion seems to occur over water or desert where there are no other objects of particular, perceived sizes. However, it is likely that modifications of the terrain would produce changes in the value of S'/θ in the visual field at the apparent position of the moon, and it is difficult to imagine a visual field which contains no values of S'/θ but still maintains the perceived distances required for the size-distance invariance type of explanation.

However, regardless of whether perceived size is to be understood in terms of perceived distance or S'/θ, the moon illusion, like Emmert's law, requires the equidistance tendency as part of its explanation. There is no reason, for example, to expect modifications of the terrain to change the perceived size of the moon unless there is some factor which makes the moon appear at some portion of the terrain. There is no reason to expect the moon to appear on the apparent dome of the sky (cf. King & Gruber, 1962), or for the horizon moon to appear at the distance of objects on the horizon, unless there is some factor to cause the moon to appear at these positions. It is asserted that this factor is the equidistance tendency. The equidistance tendency localizes the moon in depth in the visual field and consequently provides an essential condition for the operation of the processes involved in the determination of its perceived size.

The moon illusion is only one of a number of naturally occurring perceptions to which the equidistance tendency can be applied. For example, Gibson and Flock (1962) attributed an illusion concerning the underestimation of the distance of distant mountains to a reversal in the optical gradient. The equidistance tendency provides a different explanation of this illusion. The nearest mountains in a mountain range are localized perceptually with respect to the visible terrain extending from the observer. The mountains positioned behind the nearest mountains provide few cues to determine their depth position, and as a result of the equidistance tendency their apparent distance behind the nearest mountains, and thus their apparent distance from the observer, would tend to be underestimated. It seems that the perceptual consequences of the equidistance tendency must be considered in naturally occurring as well as in experimental situations.

---

[2] Boring (1943, p. 57) probably provides an illustration of this phenomenon when he notes that the apparent size of the horizon moon shrinks when viewed between the thumb and forefinger.

POSSIBLE EXPLANATIONS OF THE EQUIDISTANCE TENDENCY

The requirements which must be met by any attempt to explain the equidistance tendency are specified by the range of phenomena which demonstrate the tendency. From the point of view of the present paper, any explanation must encompass the following phenomena:

1. The equidistance tendency occurs between any visual inhomogeneities (between any contours or parts of a figure or surface).
2. The strength or effectiveness of the equidistance tendency is inversely related to directional separation (in any orientation).
3. The equidistance tendency occurs between objects viewed monocularly, or between objects viewed binocularly, or between objects some of which are viewed monocularly and some binocularly.
4. The presence of strong depth cues can reduce or eliminate the effectiveness of the equidistance tendency.
5. Under certain conditions, the equidistance tendency can modify depth perceptions resulting from cue systems such as binocular disparity and relative or familiar size.
6. The equidistance tendency can act as a resultant effect occurring between one object and a complex visual field.

Roelofs (1960, 1961) presented an explanation of the process by which monocularly observed objects appear at approximately the same apparent distance as binocularly observed objects. This explanation asserted that the image of the monocularly observed object in the one eye fuses binocularly with an unstimulated but approximately corresponding area in the other eye. Since the resulting fusional process will tend to be with respect to corresponding points, the monocularly observed object will tend to appear at the same apparent distance as the binocularly observed fixated object. However, it is difficult for the present writer to see how this explanation can be applied to cases in which the equidistance tendency occurred between objects observed either completely monocularly or completely binocularly.

Bergman and Gibson (1959) described the results of their study as indicating the occurrence of a negative aftereffect. The tendency for the slanted surface to appear in a frontoparallel plane is attributed to a perceptual shift toward a norm and is called normalization. The effect of the normalization is assumed to persist over time and results, for example, in subsequently presented frontoparallel surfaces appearing tilted in the opposite direction. It is possible, therefore, that the concept of normalization might provide a basis for understanding the equidistance tendency. There are several reasons, however, for considering this to be unlikely. It is not obvious why the effectiveness of a normalization would be inversely related to directional separation as is required by the equidistance tendency. Furthermore, since the equidistance tendency can occur in visual fields containing a large number of different objects at different distances, an explanation in terms of normalization would require the concept of a large number of competing normalization tendencies. The range of phenomena involved in the equidistance tendency is such that an explanation of the tendency in terms of any simple concept of normalization is not apt to be successful.

Several of the discussions of the equidistance tendency (Gogel, 1956b; Roelofs & Zeeman, 1957) suggest that the tendency can be treated as a factor possessing both magnitude and direction. It would seem reasonable, therefore, that the equidistance tendency can be additive or subtractive with respect to other cue tendencies in the perceptual localization of an object in apparent depth (cf. Gogel, Brune, & Inaba, 1954). Also, equidistance tendencies can subtract or add in determining a final equidistance vector. Thus, in the present paper, it is assumed that the equidistance tendency adds a depth factor to any visual field containing more than a single point of light, and that as a depth factor it enters quantitatively in competition or agreement with other depth factors in the determination of a final apparent position of an object in depth.

SOME UNSOLVED ASPECTS OF THE EQUIDISTANCE TENDENCY

It has been reported that the tendency for surfaces slanted in depth to appear in a more frontoparallel orientation changes with the amount of observation time (Bergman & Gibson, 1959; Clark, 1953; Simon, 1936; Smith, 1964). It is therefore possible that the effectiveness of the equidistance tendency (relative to whatever cue systems are available) changes with observation time. Perhaps in the absence of effective opposing cue systems or tendencies, the perceptual consequences of the equidistance tendency are rapid. This is suggested by the ability of the equidistance tendency to determine the perceived path of movement of an object attached to the rotating trapezoidal window with the entire display viewed monocularly (Gogel, 1956a). The rapidly changing directional alignments in the study with the rotating window required that the effect of the equidistance tendency be rapid in order for it to determine apparent depth. Without opposing cues or tendencies, the equidistance tendency may be quickly effective. When cues of depth are present, the relative effectiveness of the equidistance tendency (relative to that of the depth cues) possibly increases within some interval of observation time. Research is required to clarify the temporal characteristics of the equidistance phenomena both for simultaneous and successive presentations.

Additional aspects of the equidistance tendency should be considered. The notion that the equidistance tendency between two objects is a vector of a particular magnitude would suggest that increasing the number of objects at a constant distance would increase the strength of the equidistance tendency toward this distance. In a similar fashion, the strength of the equidistance tendency between objects might increase with an increase in their visual area. Also, the meaning of the term directional separation requires clarification. Is the significant directional separation to be defined physically, retinally, or perceptually? In most experimental situations a distinction between the three types of definitions would not need to be made. However, it is possible to construct experimental situations in which the three types of directional separation would not produce identical results.

REFERENCES

Beck, J., & Gibson, J. J. The relation of apparent shape to apparent slant in the perception of objects. *Journal of Experimental Psychology,* 1955, **50**, 125–133.

Bergman, R., & Gibson, J. J. The negative aftereffect of the perception of a surface slanted in the third dimension. *American Journal of Psychology,* 1959, **72**, 364–373.

Boring, E. G. The moon illusion. *American Journal of Physics,* 1943, **11**, 55–60.

Clark, C. W. Exposure time and the perception of slant. Paper read at Canadian Psychological Association, June, 1953.

Clark, C. W., Smith, A. H., & Rabe, Ausma. Retinal gradient of outline as a stimulus for slant. *Canadian Journal of Psychology,* 1955, **9**, 247–253.

Clark, C. W., Smith, A. H., & Rabe, Ausma. Retinal gradients of outline distortion and binocular disparity as stimuli for slant. *Canadian Journal of Psychology,* 1956, **10**, 77–81. (a)

Clark, C. W., Smith, A. H., & Rabe, Ausma. The interaction of surface texture, outline gradient, and ground in the perception of slant. *Canadian Journal of Psychology,* 1956, **10**, 1–8. (b)

Epstein, W., & Park, J. Examination of Gibson's psychophysical hypothesis. *Psychological Bulletin,* 1964, **62**, 180–196.

Epstein, W., Park, J., & Casey, A. The current status of the size-distance hypotheses. *Psychological Bulletin,* 1961, **58**, 491–514.

Gibson, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950. (a)

Gibson, J. J. The perception of visual surfaces. *American Journal of Psychology,* 1950, **63**, 367–384. (b)

Gibson, J. J., & Carel, W. Does motion perspective independently produce the impression of a receding surface? *Journal of Experimental Psychology,* 1952, **44**, 16–18.

Gibson, J. J., & Flock, H. The apparent distance of mountains. *American Journal of Psychology,* 1962, **75**, 501–503.

Gibson, J. J., & Gibson, Eleanor J. Continuous perspective transformations and the perception of rigid motion. *Journal of Experimental Psychology,* 1957, **54**, 129–138.

Gogel, W. C. Relative visual direction as a factor in relative distance perceptions. *Psychological Monographs,* 1956, **70**(11, Whole No. 418). (a)

Gogel, W. C. The tendency to see objects as equidistant and its inverse relation to lateral separation. *Psychological Monographs,* 1956, **70**(4, Whole No. 411). (b)

Gogel, W. C. The visual perception of size and distance. *Vision Research,* 1963, **3**, 101–120.

Gogel, W. C. The perception of depth from binocular disparity. *Journal of Experimental Psychology,* 1964, **67**, 379–386. (a)

Gogel, W. C. Size cue to visually perceived distance. *Psychological Bulletin,* 1964, **62**, 217–235. (b)

Gogel, W. C., Brune, R. L., & Inaba, K. A modification of a stereopsis adjustment by the equidistance tendency. *United States Army Medical Research Laboratory Report*, 1954, No. 157, 1–11.

Gogel, W. C., & Harker, G. S. The effectiveness of size cues to relative distance as a function of lateral visual separation. *Journal of Experimental Psychology*, 1955, **50**, 309–315.

Gruber, H. E., & Clark, W. C. Perception of slanted surfaces. *Perceptual & Motor Skills*, 1956, **6**, 97–106.

Gruber, H. E., King, W. L., & Link, S. Moon illusion: An event in imaginary space. *Science*, 1963, **139**, 750–752.

Harker, G. S. Apparent frontoparallel plane, stereoscopic correspondence, and induced cyclotorsion of the eyes. *Perceptual & Motor Skills*, 1962, **14**, 75–87.

Hartman, A. M. Effect of reduction on the relationship between apparent size and distance. *American Journal of Psychology*, 1964, **77**, 353–366.

Howard, I. P., & Templeton, W. B. The effect of steady fixation on the judgment of relative depth. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 193–203.

Judd, C. M. An optical illusion. *Psychological Review*, 1898, **5**, 286–294.

Katz, D. *The world of colour*. London: Kegan Paul, 1935.

Kaufman, L., & Rock, I. The moon illusion. *Scientific American*, 1962, **207**, 120–130. (a)

Kaufman, L., & Rock, I. The moon illusion: I. Explanation of this phenomenon was sought through the use of artificial moons seen on the sky. *Science*, 1962, **136**, 953–961. (b)

King, W. L., & Gruber, H. E. Moon illusion and Emmert's law. *Science*, 1962, **135**, 1125–1126.

Köhler, W., & Emory, D. A. Figural after-effects in the third dimension of visual space. *American Journal of Psychology*, 1947, **60**, 159–201.

McDonald, R. P., & O'Hara, Patricia T. Size-distance invariance and perceptual constancy. *American Journal of Psychology*, 1964, **77**, 276–280.

Natsoulas, T. On homogeneous retinal stimulation and the perception of depth. *Psychological Bulletin*, 1963, **60**, 385–390.

Price, G. R. On Emmert's law of apparent sizes. *Psychological Record*, 1961, **11**, 145–151.

Rock, I., & Kaufman, L. The moon illusion: II. The moon's apparent size is a function of the presence or absence of terrain. *Science*, 1962, **136**, 1023–1031.

Rock, I., & McDermott, W. The perception of visual angle. *Acta Psychologica*, 1964, **22**, 119–134.

Roelofs, C. O. Convergence and depth perception. *Acta Psychologica*, 1960, **17**, 193–217.

Roelofs, C. O. Equidistance tendency. *Acta Psychologica*, 1961, **18**, 368–382.

Roelofs, C. O., & Zeeman, W. P. C. Apparent size and apparent distance in binocular and monocular vision. *Ophthalmologica*, 1957, **133**, 188–204.

Simon, M. Über egozentrische Lokalisation. *Psychologische Forschung*, 1936, **21**, 113–131.

Smith, A. H. Gradients of outline convergence and distortion as stimuli for slant. *Canadian Journal of Psychology*, 1956, **10**, 211–218.

Smith, A. H. Outline convergence versus closure in the perception of slant. *Perceptual & Motor Skills*, 1959, **9**, 259–266.

Smith, A. H. Delay of response in the perception of slant and shape. *Canadian Journal of Psychology*, 1964, **18**, 235–244. (a)

Smith, A. H. Judgment of slant with constant outline convergence and variable surface texture gradient. *Perceptual & Motor Skills*, 1964, **18**, 869–875. (b)

Smith, A. H. Relation of slant and shape judgments in reduced viewing. *Canadian Journal of Psychology*, 1964, **18**, 224–234. (c)

Smith, O. W., & Smith, Patricia C. Interaction of the effects of cues involved in judgments of curvature. *American Journal of Psychology*, 1957, **70**, 361–375.

Smith, Patricia C., & Smith, O. W. Veridical perceptions of cylindricality: A problem of depth discrimination and object identification. *Journal of Experimental Psychology*, 1961, **62**, 145–152.

Stavrianos, Bertha K. The relation of shape perception to explicit judgments of inclination. *Archives of Psychology*, 1945, **41**, 1–94.

Werner, H. Dynamics in binocular depth perception. *Psychological Monographs*, 1937, **49**(2, Whole No. 218).

Werner, H. Binocular depth contrast and the conditions of the binocular field. *American Journal of Psychology*, 1938, **51**, 489–497.

(Received March 15, 1965)

Psychological Bulletin
1965, Vol. 64, No. 3, 164–182

TOWARD A THEORY OF OPEN AND CLOSED GROUPS

ROBERT C. ZILLER
University of Delaware

4 characteristics differentiate groups in which membership is in a constant state of flux (open groups) as opposed to groups in which the membership is relatively stable (closed groups): time perspective, equilibrium, frame of reference, and changing group membership. The significance of these characteristics for social behavior was explored. A number of tested and testable propositions concerning group stability and social behavior emerged from a rapprochement of research, relevant subtheories, and concepts associated with open- and closed-group behavior. The failure to consider the dimension of group stability in most previous social-psychological research poses a question concerning the generality of social-psychological theories which are based upon research which ignores the pervasive dimension of group stability.

It has long been recognized by organization theorists (Hemphill, 1950) that group stability is an essential construct with regard to group behavior. However, even a cursory investigation of research reports concerning social behavior reveals that social scientists have been largely concerned with groups whose membership remains constant. This traditional model of social research is scarcely isomorphic with regard to social systems as we would know them if we did not ignore or repress the minatory forces of illness, death, and changes in group composition in general.

Certainly a basic characteristic of social systems is that they maintain themselves in a state of perpetual change of components. It is true that within a given time span the composition of the social system appears static. Perhaps we are deceived by our own parallel movement along the time flow and the relatively steady state of the system at the molar level.

Although Frenkel-Brunswik (1949) and Heider (1958, pp. 173–194) hypothesized individual differences in response to change in general, theory and research with regard to a general response to changes in group composition are equivocal. The equivocation probably evolves from the novelty effects inherent in the change condition involving moderate stimulus intensity, and the hypothesis (Berlyne, 1960) that human beings normally strive to maintain an intermediate amount of arousal potential. Although there have been few attempts to generalize Berlyne's hypothesis to conditions of group member homeostasis, the hypothesis suggests higher member satisfaction is associated with a modicum of membership change. Although the intensity of the membership change stimulus is difficult to define, it will be assumed here that membership change usually results in member high arousal potential. Assuming high stimulus intensity under conditions of membership change, it is proposed that in approaching a dynamic environment involving an array o familiar and unfamiliar social objects, the tension level of a cognitive organism tends to increase as a consequence of negative reinforcement in previously encountered unstructured situations. In addition, an unfamilia situation poses a threat to individual security. A confrontation of the self may deman a reevaluation of the self, which, in turn demands a redistribution of activity and additional energy expenditure. Thus, March an Simon (1958) propose that "in general an up to a fairly extreme point, increased predictability yields increased satisfaction for mos people [p. 94]." Becker (1950, p. 52) supports this latter viewpoint. His proposition described the hostility which attends the advent of a newcomer to relatively stable groups since the newcomer represents the forces o change.

On the other hand, at the common sens

---

[1] This investigation was supported in part by Air Force Contract Number AFOSR 62-95, United States Air Force, Air Force Office of Scientific Research, Office of Aerospace Research, Washington 25, D. C.

level of psychology, a number of assumptions have been made concerning the positive aspects of membership change. It is said, for example, that newly commissioned officers in military units in the 1930 decade were reassigned frequently in order to avoid constant reminders of initial failures. Highly capable coaches and members of professional baseball teams with disappointing records are replaced, and the need for "new blood" or a new environment is offered in mythical explanation. Speculation is widespread as to the number of years that a given group can profitably or safely retain the same leader (McVey & Hughes, 1952). Finally, changes in personnel are the bane of embezzlers. In general, these and other speculative change strategies in group membership derive from an implicit theory of open-closed groups.

Perhaps the first systematic analysis concerning the open-closed group phenomenon is found in the writings of Simmel (Wolff, 1950). Specifically, Simmel speculated extensively concerning the unique characteristics of the stranger in an otherwise stable community. The nucleus of the characteristics discussed is composed of the concepts of mobility and objectivity. "The fundamentally mobile person comes in contact, at one time or another, with every individual, but is not organically connected, through established ties of kinship, locality, and occupation, with any single one [p. 404]." Moreover, the stranger "is not radically committed to the unique ingredients and peculiar tendencies of the group, and therefore approaches them with the specific attitude of 'objectivity' [p. 404]."

Though Simmel was concerned largely with the characteristics of the stranger, the arguments also have implications for groups which do or do not have contact with strangers, that is, open and closed groups. The open-closed group concept derives rather broadly and only analogously from a general systems theory framework (Bertalanffy, 1950). A basic unit of this approach is a "system" defined as a "set of elements standing in interaction [p. 23]." Systems in this very general sense may be categorized as either open or closed. A closed system is isolated from the environment; new elements are excluded and the departure of elements is precluded. It is proposed that most living organisms are initially open systems, since they are maintained "in a continuous inflow and outflow, building up and breaking down of components, neither being, as long as it is alive, in a state of . . . equilibrium but maintained in a so-called steady state . . . [Bertalanffy, p. 23]."

Extending this framework to social systems, an open group may be defined as an interacting set of persons in a continuous state of membership flux; for example, an Army unit under combat conditions. Under these conditions of social metabolism, the elements of the complex are temporarily interrelated owing to periodic or frequent but unscheduled replacements, removals, or additions of elements. In closed groups, on the other hand, the elemental composition remains constant. Polaris submarine crews, for example, remain submerged for three months without even radio contact with the outside world.

Four fundamental characteristics appear to distinguish open groups from closed groups: a reduced time perspective; the achievement and maintenance of steady states (equilibrium); an expanded frame of reference; and, of course, changing group membership. Under the heading of changing group membership, we shall be concerned with the topics of mobility and power, assimilation process of the newcomer, and leadership change.

TIME PERSPECTIVE

By definition the lives of the members of closed groups are inextricably intermingled for a given period of time. Conversely, members of open groups are aware of the transitory nature of these relationships. The future is indistinct; the present dominates. Implementation of group decisions tends to be activated with greater haste. If the present membership is to benefit from a given group action, the action must be undertaken in the immediate future. Postponement may be tantamount to inaction for those members who may be removed or replaced.

Sorokin (1927) states that contemporary Western society demands a reduced sensitivity to the environment. He contends (p. 518) that our nervous system would be wrecked completely if it were sensitive to the innumerable phenomena which surround us in our permanent mobility. At least two alternatives are possible, group orientation or self-orientation.

Group structure growing from the members' attachment to the group as a group rather than to unique relationships with individual members may insure a degree of stability in an otherwise changing social environment. For example, in order to insure the maintenance of the norm of reciprocity within an open group, loyalty to the group must be assumed. Reciprocity is defined (Gouldner, 1960) as the "mutually contingent exchange of gratification," and is assumed to be a universal norm. Minimally, reciprocity demands that: (*a*) people should help those who have helped them and (*b*) people should not injure those who have helped them. In open groups, however, a system of sanctions may be difficult to maintain. Thus a beneficent member of the group may fail to be recompensed by another member who departs before compensation is possible. It is proposed that in open groups interpersonal interdependency surrenders to person-group interdependency. Members bestow favors on other members in the name of the group, and the group of the larger organization of which it is a part is obligated to assume the responsibility of reciprocation. Herein lies the concept of loyalty and its rewards. In closed groups, then, the reciprocity norm functions among individuals; in open groups, the reciprocity norm functions between the individual and the group as a group.

Self-orientation may also provide a relatively stable basis in an otherwise changing social field. Thus, in anticipation of being forced to change groups, the members may become primarily concerned with transferrable acquisitions, that is, returns or benefits advantageous primarily to the single mobile member.

To counteract the possibility of this short-term perspective, the group may invoke the concept of a universal group or a universal philosophy or religion to which all men belong or adhere. In this way, intergroup mobility becomes intragroup mobility and a degree of stability is achieved by the system.

Consideration of time perspective in open and closed groups has implications for formal game theory. Thus, it is proposed that a player does not select a strategy to maximize individual gains as readily in closed groups as in open groups since extra-game costs of an interpersonal nature may be involved.

EQUILIBRIUM

The theoretical framework which associates open-closed groups and equilibrium was sketched by Chappel and Coon (1942).

> A state of equilibrium may be defined as follows: if a small force is impressed upon a system, a change or adjustment takes place within the system, and once that force is removed, the system returns to approximately its previous state [p. 43].

One such force is a change in group membership.

It has been repeatedly suggested that instability is a basic shortcoming of an open group system. Changing leadership at an inopportune time, a rapid turnover of personnel, a sudden exodus of members, a loss of group identity, a concern for sources of replacements or additions, and the unpredictability of future social contacts are just a few of the problems inherently involved in open groups. It is assumed here that a constant state of disequilibrium is nonfunctional. Thus, it is hypothesized that the members of open groups develop procedures which minimize the disruptive effects of membership changes and establish a state of dynamic equilibrium through considerations of group size, homogeneity of personnel, and turnover rate.

Group Size

It is proposed that group size is a more critical variable in open groups than in closed groups, particularly with regard to very small groups. If the group is composed of only two persons, the departure of one of these dissolves the group. Even with three members, however, there remains a substantial probability that two members may leave simultaneously, particularly in view of a possible concatenation effect. Since the entire organization is more threatened by the loss of a member from a small group than a large group, the members of small groups may overreact to the threat of group disintegration and leave the group prematurely. Moreover, under these conditions, undue power is inherently accorded the mobile individual, since he possesses the power to destroy the group. Group size is a critical consideration in terms of group stability from yet another point of view. The probability of compatibility between the new member with at least one other member is increased with an increase in group size. Thus, the assimilation process may be presumed to be facilitated by increasing the number in the host group, thereby increasing the likelihood of the newcomer remaining in the group. If, however, the membership is homogeneous, the effects of group size are attenuated.

Increased membership heterogeneity, as with increased group size, is presumed to increase the probability of matching members and thereby increase compatibility and group stability under open group conditions. A heterogeneous grouping of personnel also facilitates group stability by providing a more extensive pool of replacements which insures an adequate flow of personnel into the unit.

Turnover Rate

Trow (1960) assumes that "turnover necessarily disrupts informal understandings and working arrangements" and therefore affects the coordinates of the group's activities and discussions. Moreover, turnover usually entails a loss of part of the group's "memory."

Increasing the rate of turnover presumably may reach the height where personnel problems hold precedence over productivity. Open groups, then, are required to conduct membership changes at a rate which is minimally disruptive. One means is to make changes in a block, rather than extend the transition period (Trow, 1961). However, under military combat conditions, the replacement of a large block of personnel may interfere with the confidence of the remainder of the organization so as to interfere with its effectiveness under stress (Anonymous, 1946). The optimum rate of turnover appears to be a function of the task demands.

A relationship between turnover rate and membership satisfaction was reported by Darley, Gross, and Martin (1951). Satisfaction of girls with their own dormitory life was found to be related negatively to the stability of the dormitory population.

Finally, a relationship between growth rate and institutional strength is hypothesized by Chapin (1957). On the basis of a comparison of eight old and eight young churches, it was hypothesized that weak institutions are a result of disintegration from burgeoning group size or lack of new blood.

It is also assumed that one of the dangers of membership change is oscillation due to an overreaction to change. In order to counteract this tendency and to maintain efficiency during readjustment, open systems may develop a lag or decrease in sensitivity to change (Herbst, 1954, p. 333). Lag is the amount of time required before changes in one component are communicated or transferred to other components. One concrete means of building a lag into the system is to suggest that the host group members react to the position to which the new member has been assigned rather than to his characteristics as a unique individual. Again, military replacements are an excellent example, since the position of the individual in these units is clearly depicted with appropriate symbols prominently displayed on the uniform.

One of the inherent problems of open groups is the maintenance of a flow of personnel in and out of the system, yet maintaining the steady state of the system. Here, age heterogeneity is necessary in order to avoid mass replacements as the members advance in age.

Assistance in making a successful and satisfying transition to another group also contributes to group stability by reducing the tendency for former members to derogate the donor group (Caplow & McGee, 1958).

Another empirical development for maintaining group stability among open groups such as colleges, military service units, or political groups is the organization of former members as in alumni clubs, veterans' organizations, political parties, and perhaps the Actors Guild. The functions of these latter more stable organizations are manifold but include maintaining the flow of personnel in and out of the organization which they serve in an auxiliary capacity and, perhaps, acting as an extended reward system within which the parent organization can exercise sanctions toward a transgressor who has left the organization by precluding membership in the graduate organization. With these graduate groups, then, a former member, in effect, can never escape the influence of the parent organization. In this respect, the open group assumes some of the characteristics of a closed group. Groups in which careers are short-lived or hazardous appear to demand the services of the extended group as a means of system stability.

FRAME OF REFERENCE

The term "frame of reference" commonly refers to the ground against which the figure is perceived (Newcomb, 1950, p. 94). We shall use the broader definition of the term which includes all the factors, objective and subjective, which influence the perception of the figure.

It is proposed that open groups, in contrast to closed groups, have an expanded frame of reference (Redfield, 1947; Sorokin, 1927). These sociologists aver that members of an immobile society learn, as a rule, only a definite course of ideas, opinions, beliefs, and values. They have little opportunity to listen to many ideologies and learn different beliefs; and, as a consequence, their belief systems are inflexible.

Several advantages inhere in the use of open groups or, in effect, a larger population as a frame of reference for status evaluations rather than a fortuitiously selected small closed group. In the expanded frame of reference of the open groups in which transfers frequently occur, more accurate and more reliable ratings of the members are possible, since all members tend to be exposed to paired comparisons with a wide variety of others. Similarly, more accurate self-concepts are encouraged within open systems. Since it is less possible to insulate and isolate oneself, inflated self-estimates do not remain unchecked or unchallenged for long; nor are erroneous self estimates buttressed for long by peculiar status achievements within a given small social unit.

These same arguments may be applied to groups. Thus, in open group systems, islands of cultural or organizational lag are less probable than in closed groups (see Redfield, 1953). The proposition concerning an expanded frame of reference in open groups also has implications for group creativity and change in group set.

Creativity

Many social scientists have long held that creativity flourishes in societies where there is a continuous flow of new exogenous members (open societies), and is curbed in societies from which strangers are restricted. The usual rationale is that isolation results in highly specialized conceptual frameworks and rigid limitations on the kind of behavior deemed acceptable in a given situation (Redfield, 1947). Sorokin (1927), however, proposed that in horizontally immobile societies with their "permanent and monotonous environment, there is little incentive for invention" and no favorable combination of conditions which may suggest the ideas of invention.

Empirical evidence supporting this hypothesis involves research teams in industrial laboratories (Shepard, 1955). The results suggest that group productivity, creativity, and enthusiasms are negatively related to length of association of the group members. In a laboratory study in which the creativity task involved the composition of captions for a *Saturday Evening Post* cartoon, groups which experienced membership changes (open groups) were more creative than stable groups (closed groups) (Ziller, Behringer, & Goodchilds, 1962).

The latter results were interpreted in terms of social facilitation (Allport, 1924). The performances of the host and donor groups may be compared by the newcomer, whereas the judgments of the closed group's performance is necessarily absolute. Thus intergroup comparison or competition may be the intermediate variable relating changing group membership and creativity.

Change in Group Set

It is hypothesized that a change in group composition facilitates a change in the group set toward group outcomes. For example, it is proposed that a failure atmosphere or attitudes constituting low group evaluation following group failure are dissociated more readily if membership is altered. Furthermore, it may be proposed that the disruption of group continuity provides a convenient basis for dividing the group history into two time segments—the past and the future—and for compartmentalizing failure with the past. Indeed, following failure, a change of any kind may provide a plausible basis for dissociating the past and perceiving the future as a "new deal" or a new group era relatively free from obtrusive reminders of past unpleasantries (New Year's Eve phenomenon). Change per se is efficacious. Any one of a variety of changes may provide a rallying point or a convenient defense mechanism which serves a therapeutic function for groups embarrassed by memories of failure.

On the other hand, it may be proposed that groups with a history of success attempt to minimize change in order to preserve the identity of the group and the members' perceptions of the nebulous but accepted "winning combination." It is difficult for group members as well as for professional social scientists to define the basis for group success. Thus, in the event of group success, it is predicted that the group members tend to adhere in every detail to the previously successful pattern of group behavior. Since this is impossible under conditions of membership change, a change may be perceived as potentially corrupting or as a threat to the group's rather imponderable but, nevertheless, successful pattern of behavior.

Changing Group Membership

The most obvious characteristic that distinguishes open groups from closed groups is the inevitable change in group composition. In open groups, members may be added, removed, or replaced. Replacement, however, is a composite of a removal and an addition so it need not be discussed separately.

Changes in group composition may be strategic or merely in response to the exigencies of group life. We shall be primarily concerned with planned change. The entire section is introduced by a discussion of power, a basic consideration in relation to planned changes in group composition.

Power

Lewin (1951) has defined the power of *b* over *a* as the maximum force that *b* can put on *a* toward some region relative to the maximum resistance that *a* can mobilize against *b*'s force. Cartwright (1953) has used this concept to describe some of the consequences of power in relatively permanent groups or, as used here, in closed groups. He notes that a higher degree of power of *b* over *a* carries potential, whether or not *b* has ever used his power. In closed groups, *b*'s power may be exercised either in the present or in the future.

Thus, in closed groups, any act between *a* and *b* potentially has consequences for the power relationship between *a* and *b* in the future. On the other hand, in open groups, either *a* or *b* may leave the group voluntarily. This contingency tends to reduce the power aspects of the relationship since the length of time is reduced over which either person is able to impose sanctions on the other. The person toward whom sanctions are being directed may be removed from the group and the other person's sphere of influence; thereby, in Lewin's terms, increasing the resistance that the person can mobilize against the other's force by withdrawing from the field. Of course, to counteract the possible reduction in length of time in which something may be exacted in open groups, the sanctions may become more severe and immediate in open groups.

Power and the Newcomer

Some particularly unique power considerations arise with regard to the newcomer and imminent egressor under open-group conditions. By definition, the newcomer has recently been a member of another group, often a group quite similar to the host group. At the very least, then, the newcomer possesses a more extended frame of reference than those members of the host group who have worked exclusively within the boundaries of the host group. In terms of interpersonal evaluation, the newcomer is in a position to make intragroup and intergroup comparisons among the group members. Similarly, the newcomer is in a position to render invidious comparisons between the host and donor groups; but, of course, the group is in a similar position if the member is a replacement (Gouldner, 1954).

Various subgroups among the main body may become concerned as to in which subgroup the newcomer will cast his lot, for by becoming aligned with one or the other of the various informal and sometimes competitive or even hostile subgroups of an organization, the equilibrium of forces within the host group may be threatened. Thus, the allegiance of the newcomer may be actively solicited by each of the subgroups, and the newcomer, through his position as an uncommitted member and perhaps a pivotal member, is accorded a strategic power role in the group (Snyder, 1958).

Moreover, it has been demonstrated (Wright, 1942) that the newcomer tends to be regarded as a guest rather than as a regular member during the initial phase of adjustment. The prerogatives of the guest being to some degree inviolable, the newcomer is in a position where the regular members may find it necessary to defer to the newcomer's wishes. Should the "guest" depart unexpectedly soon after his arrival, the host group would be somewhat embarrassed publicly.

It must also be noted that in the early phases of the new relationship the newcomer is often not totally committed either financially or psychologically to the new group. Therefore, the newcomer is more mobile in a variety of ways. Mobility, it will be shown, is related to power.

Once the newcomer has been assimilated and categorized with respect to some of the major dimensions of personality or subgroups in the organization, the newcomer loses a degree of power associated with individualistic behavior. It is assumed here that a degree of power resides in the role of the stranger, if only by virtue of the fact that individuals can predict the stranger's behavior with a minimum of certainty. Reference is made here to game theory. The stranger in these terms is resorting, so far as the host group members are concerned, to relatively random behavior. The older members are themselves more predictable and, in fact, are constrained to be in relation to their position and role in the group. Thus, the regular members are more vulnerable to various game strategies than the newcomer.

In general, some of the power advantages of the newcomer have been outlined. Conversely, various negative power components are inherent in the role: lack of knowledge of the informal group structure, cautious behavior to preclude premature and erroneous evaluations, unfamiliarity with group norms, and inability to check the validity of personnel evaluations with other members.

Power and the Egressor

Another role peculiar to open groups is that of the egressor, the member whose departure is imminent. In contrast to the newcomer, it is hypothesized that the power of the egressor diminishes directly in relation to the remaining membership time. Group tenure and power are directly related. In Lewin's (1951) terms again, since the egressor (E) will depart on a given date, the group's ability to resist forces emanating from E may be maximized through certain strategies. For example, the group may delay reaching a decision or acting upon a decision until after E's departure. E may even be omitted from the decision-making group.

By the same token, however, E can resist forces emanating from the group. Thus, in military circles the phenomenon of reduced motivation to action or increased resistance to group pressure as the enlistment period draws to an end is referred to as "stacking arms" or "going casual." Moreover, E is a potential threat to the donor organization. Since his loyalties are indeterminate, a marginal man as it were, E may be free to reveal negative aspects of the donor group without fear of sanctions.

Although these propositions are largely concerned with individual group members, they may also be extended to the groups as groups. For example, a small manufacturing unit may encounter fewer problems of labor-management relations since the parent organization may abandon the operation at a lower cost. Here mobility of the manufacturing plant presents a departure threat to the bargaining labor group which may be less mobile for a variety of reasons to be discussed later. Thus again mobility may enhance power.

The power of E is contingent upon at least two other qualifications: (a) the characteristics of the group which he is expected to join and (b) the probability of a replacement.

If E is to be replaced, it behooves the group to react to E in terms of the position occupied since adjustment problems during the transition period will be minimized if the position remains active. If, on the other hand, the position is to be eliminated from the organizational structure, E's power will be minimized during the final phases of transition.

The nature of E's host groups must also be considered in relation to E's penultimate power. It may be hypothesized that E's power is directly related to the power of the host group in relation to the donor group. The rapid diminution of the power of the member approaching retirement is relevant. Since the retiring member usually does not rejoin an organized social unit, power derived from organizational affiliation is withdrawn. On the other hand, if E is to join another group whose area of influence extends to the donor group, E's power probably does not decline to the same degree or rate. In this regard, it would be helpful to deduce the power path of leaders and other group members as they approach retirement. For example, assuming the power path decelerates, at what point is the change first observed by the members and E, and what is the nature of the deceleration curve?

In a very general sense, then, it is proposed that future mobility and past mobility are related to the power of an individual in the group. The nature of the power is qualified by some of the conditions discussed here.

The Departure Alternative in Open Groups

As has already been indicated, a fundamental characteristic of open groups is that a flow of members into and from the core of the complex is expected, happens, or is possible. It is the last qualification with which we are concerned in this section.

At the gross level of abstraction, organizations may be differentiated with respect to the ease of exit. In a sense, this is merely a special case or emphasis of the general open group classification; but a number of ramifications emerge which are relatively specific to this attribute.

Organizations which are relatively difficult to leave are readily recalled: a prison, mental institutions in some cases, hospitals, military groups, an island community or isolated community in general, and perhaps some nations. What, now, are some of the social-psychological consequences of barriers to departure or ease of departure?

Here again, we will invoke a decision-making framework. In general terms, the option to leave the group presents an expanded range of alternatives to the decision-maker; the individual can withdraw membership. While relatively little research has been concerned with the range of alternatives in a decision-making situation a number of hypotheses are suggested.

As the number of alternatives are increased, in a decision-making situation, the degree of conflict is intensified. Thus, the introduction of the departure alternative increases the degree of conflict for the group members.

March and Simon (1958) have speculated that "the greater the number of perceived alternatives to participation available in the external environment, the less important the consequences associated with variations in conformity to organizational demands [p. 58]." For example, they hypothesize that the greater the number of unemployed workers the fewer the perceived alternatives to participation.

It has also been hypothesized (Ziller, 1955) that an expanded range of alternatives on scale of judgment is associated more frequently with a decision of superior quality. The rationale for this proposition rests on the assumption that the probability is greater that the superior decision will be included in a relatively heterogeneous and expanded range of alternatives. In terms of the decision of group membership, it may be hypothesized that an increase in the number of alternatives available to the individual will lead to a superior decision and perhaps to a more personally satisfactory decision.

Four variations in the initiation of change must be considered: (*a*) only the individual member may initiate change in group membership and group approval is not required, (*b*) only the group may initiate change proceedings for the individual, (*c*) either *a* or *b*, (*d*) both the individual and the group must agree on change proceedings. It is immediately recognized that these four conditions describe the focus of power for initiating membership change. Condition *d* probably requires an arbiter.

Given Condition *a*, the individual member may be expected to retain group membership as long as his needs are satisfied to a greater degree by that group than by any other group in which membership is available. The member is in a buyer's market, as it were. Moreover, it is difficult to impose sanctions on group members under these conditions. Assuming that membership in similar groups is available, the individual is free to withdraw membership or retain membership but withdraw from active group participation. Moreover, if additions are precluded and replacement is the only possible means of membership change, the power of the members under these conditions is enhanced.

The power of the individual under Condition *a* derives from several sources. If the individual has assumed a position and role of responsibility in the group, the group has become, by definition, dependent upon this member, to a degree. Thus, if a responsible member of the group considers removing his skills from the group, the state of equilibrium achieved by the group is threatened. Of course, the loss to the group is in proportion to the difficulty of finding a replacement. However, the longer a member remains in the group, the greater the group's dependence on the member, if only in the way of group memory or the idiosyncratic interdependencies of the group members.

The departure of a single member tends to disturb the equilibrium of the group. Now, during the period of transition prior to achieving equilibrium at a new level, the departure of other members may be precipitated. In the eyes of a potential egressor the difference between the difficulties of transferring membership to another group and adjusting to changes in membership in the immediate group are less than the differences between the difficulties of transferring membership to another group and adjustment in a relatively stabile group.

Finally, the power of the members of groups under Condition *a* derives from the threat to the group image. When a member leaves under Condition *a*, the departing member communicates, in effect, to the other members and other potential members that the group did not satisfy his needs and, by generalization, may not satisfy the needs of other members to the same extent as other available groups.

Under the Condition *b* of membership change, the group may be said to retain each member for that period of time that the individual satisfies the needs of the group better than any other potential member and at the lowest cost to the group. Here, then, the group is in a buyer's market; departure under Condition *b* connotes individual inadequacy in contrast to Condition *a* which connotes group inadequacy. Thus, powerful sanctions are at the disposal of the group, and extreme pressure toward conformity may be expected.

This condition may be observed in cultures where only the male can initiate divorce proceedings, in a military setting during periods of conscription, or, in general, under circumstances where the group holds precedence over the individual.

Under Condition *a*, the cost of departure is a primary concern of the group since the individual may be assumed to weigh this cost in his decision to leave. Similarly, under Condition *b*, the cost of departure is a primary concern of the individual. Under Condition *c*, a game theory model obtains. The cost and rewards to the group and individual are under constant assessment; and it may be predicted that a considerable amount of energy is expended by the parties to assess their relative positions in order to remain in a bargaining position. Neither the group nor the individual can afford to relax their continuous surveillance procedures; the other may be probing for an advantage.

Under Condition *d* the game theory model still applies, but the decision in terms of costs and rewards is usually arrived at by means of arbitration with at least a third party. By this means, individual power strategies are circumvented or at least muted.

Each of these conditions for departure also is related to conditions for joining the group. For example, it is eminently probable that the decision to join a group under Departure Condition *b* is entered into with greater deliberation than under Departure Condition *a* since a greater degree of personal commitment is involved.

Similarly, these same departure conditions are related to techniques of personnel selection. A group which can discard personnel with relative impunity and minimum difficulty is less likely to be concerned with selection techniques. The on-the-job situation provides a convenient selection criterion under some circumstances.

The decision of a member to leave the group in favor of another group may have an extensive impact on the group from yet three other points of view. As is indicated by the opening phrase, *favor of*, the decision to leave the group of origin may be perceived as unflattering to the donor group and flattering to the host group (Caplow & McGee, 1958).

However, when a colleague leaves the group it inherently enlarges the sphere of the members of the original group to include some members of the emigree's new group. The implicit derogation of the donor group through departure and the enlargement of the social world may tend to activate other departures from the initial group, particularly if several members leave simultaneously thereby creating an impression of down mobility in the original group.

Finally, the departure of a member, and particularly the leader, may precipitate a concatentation of changes in the organization since the organization may be less reluctant to institute changes. Ideas and complaints long dormant in the minds of members and administration may now find expression. Since a reorganization is required in any case with the changing group membership, a thorough overhaul may be facilitated.

Strategy of Personnel Changes

Problems associated with changing group membership have probably been recognized for centuries. The adoption of the dynasty rule and primogeniture principle to the succession of kings was undoubtedly a response to an awareness of succession difficulties.

Machiavelli (circa 1532), in *The Prince*, recognized the problem and described a stratagem whereby the Prince succeeded to the throne of a newly conquered kingdom, enforced drastic changes in government, and yet won the everlasting loyalty and gratitude of the subjugated people. To accomplish this, a myrmidon was first assigned as ruler by the Prince. This subordinate effected with dispatch and utter ruthlessness the changes deemed necessary by the Prince. When all had been accomplished to the Prince's satisfaction, the Prince entered the city, convoked the people along with the myrmidon, and slew the scoundrel before all as an act of justice and atonement—a social gambit, as it were.

The ubiquitous nature of the personnel change problem finds expression in the gnome, "Never change horses in the middle of a stream." Finally, an old English sage has advised, "Never marry a widow unless her husband was hanged." It is curious to note that the latter bit of wisdom is supported by Gouldner (1954) who refers to the tendency to compare the replacement unfavorably with the departed member as the "Rebecca Myth." The tendency for widows to exaggerate the virility of their late husbands (Bernard, 1956) may also be cited in support of this last hypothesis.

Thus a more systematic analysis of some of these ancient bits of wisdom may prove productive. The dynasty and primogeniture methods of succession immediately suggest, for example, a decision-making model in which an automatic decision-making device is used. By imposing an automatic decision-making device, internecine affairs are avoided, and the state is not vulnerable during an unduly protracted period of interregnum. Thus, in contrast to a dictatorial form of government, democracies have institutionalized an automatic succession procedure which presumably insures a more stable organization over time (Brezezinski & Huntington, 1964, pp. 89–109). Other automatic succession procedures include: promotion in accord with longevity or age, the exclusion from consideration of members of the group because of such visible characteristics as sex or race, and random selection by lots.

As was suggested earlier, change per se is hypothesized to be disliked by some individuals under most circumstances. In addition to the explanations proffered earlier, speculatively, a change in personnel may be perceived unfavorably in general because it reminds one that everyone, including self, is replaceable. The generalized antichange disposition may be displaced toward the newcomer, the characteristics of the newcomer notwithstanding. However, if the replacement is delayed (a period of mourning, as it were, in which it is taboo to replace the departed member) the recognition of the need for a replacement may surmount the antipathy toward changes in personnel and counteract the so-called "Rebecca Myth" phenomenon.

In general, it is hypothesized that a newcomer will be assimilated more readily by groups in which there is a recognized need for an additional member. It may be assumed, for example, that members of successful groups, in contrast to those in groups that have experienced failure, perceive their group as a self-sufficient, self-contained, and complete social unit. The introduction of a new member into such a group may precipitate examination of, and changes in the organization and structure of the group, the methods of sharing group rewards, and the group's modes of adjustment to its environment. Thus, to a successful group, the newcomer represents a potentially disruptive force threatening to stimulate changes in positively reinforced behavior. Unsuccessful groups, on the other hand, may be expected to perceive the newcomer as a possible resource and as an agent or catalyst of change in a situation where change is needed (Ziller & Behringer, 1960).

It will be assumed that there is a tendency to compare the newcomer invidiously with the the former member. It is also quite conceivable that if the newcomer is more favorably treated than the departed member, the assimilation process may be further retarded and the change situation further exacerbated. In open groups the preferential treatment of the latecomer may be interpreted as evidence that a premium is placed on mobility rather than loyalty.

Still, in some respects, the replacement's role is preferable to that of the addition: The replacement occupies a position vacated by his predecessor, a position to which communication channels have been previously established. For the addition, the position must be defined and established.

More simply, perhaps, the replacement represents a less abrupt change both for the newcomer and the group. The position assignment permits the new member to perform various group functions and accumulate idiosyncratic credits (Hollander, 1958) which allow for greater individual adjustment following the initial orientation phases of the assimilation process. Stated otherwise, the position assignment enables the newcomer to focus on the task demands without unduly complicating the situation with a confusing array of interpersonal considerations which can only be resolved over time.

Membership transfer may also be facilitated by avoiding an abrupt exodus. This may involve the provision of time to complete ongoing activities or transfer responsibility for these activities to other members of the donor group. A period of "independence training" may also be necessary.

The abruptness of change may also be palliated by increasing the number of contacts between the prospective new member and the host group and simultaneously decreasing the number of contacts with the donor group. This technique has been advocated by social case workers in an adoption situation where the child is initially under the care of a foster parent (Gerurd & Dukette, 1954).

Similarly, in order to reduce the abruptness of change ("cultural shock") from institutional life to civilian life, "half-way" houses have been developed for persons who have been discharged from penal institutions. In this interim period, the newcomer to the community has the support of a number of peers with similar problems.

The use of interim groups to facilitate transfer between groups provides opportunities for exchange of ideas concerning mutual problems and the discharge of anxieties through catharsis. However, prolonged attachment to this mediating group may also delay the assimilation process.

A variation of the interim group device for facilitating assimilation or transfer is the assignment of two members simultaneously into the same host group. However, if the number of newcomers in relation to the number of members in the host group approaches a substantial percent of the host group, a disintegrating effect may displace the planned assimilation effect (Anonymous, 1946). The determination of the optimum ratio of new members to regular members is amenable to empirical test.

The use of sectarian groups to cushion the "cultural shock" of rural migrants to urbanism was studied by Dynes (1956). Using attitude measures, it was found that the sect does cushion the shock if the migrant is from a lower socioeconomic level.

It has recently been demonstrated (Aronson & Mills, 1959) that the difficulty of achieving group membership is directly related to the perceived value of the membership. This relationship is explained in relationship to Festinger's (1957) theory of dissonance-consonance in a decision-making situation. It is assumed that there are unattractive as well as attractive aspects of membership in any group. If a person has undergone an unpleasant initiation his cognition that he has subjected himself to an unpleasant experience in order to achieve group membership is dissonant with the awareness that some aspects of the group are unattractive. Consonance can be effected by either minimizing the initiation or overemphasizing the attraction of the group. As the severity of the initiation increases, it becomes increasingly difficult to ignore or minimize the negative aspects of induction, leaving only the alternative to exaggerate the value of the group.

It is also hypothesized that in open groups, in contrast to closed groups, a change in membership is not likely to assume undue importance or need not be perceived as readily as a calculated reorganization of manpower. Thus, strategic membership changes can be made with greater ease in groups which experience membership changes routinely. For example, a sorely needed change in leadership need not prove as embarrassing to either the leader or the group under conditions where changes in group membership are regular occurrences.

Assimilation of the Newcomer in the Group

Various aspects of the process of the assimilation of the new member were discussed under the topics of "Power" and "Strategy Effect in Personnel Changes." Here the assimilation process will be considered in detail.

Assimilation is defined here as the process by which a newcomer establishes a positive affective identification with the host group; and the host group, in turn, comes to perceive the new member as an integral part of the social unit. For more than forty years the theoretical approaches to the newcomer phenomenon have been based largely on the writings of Simmel (Wolff, 1950). In general, Simmel described the advantages of a stranger in a comparatively stable society. The stranger's or "potential wanderer's" advantages were said to derive from the following characteristics of his position in the group: mobility, objectivity, confidance, freedom from convention, and abstract relations. With regard to mobility it was said that the stranger may interact with the members of the group but is not intrinsically related to any member through kinship, locality, or occupation.

Moreover, it is more probable that a stranger can adapt an attitude of objectivity with regard to group practices since he is not committed to the pressures toward group conformity. In support Simmel cited the practice of ancient Italian cities to appoint judges from outside their realm because endogenous citizens were not free from family entanglements and party interests. Related to the objective role is the observation that strangers (especially the stranger *passant*) often are trusted with confidences which are withheld from more closely related stable members of the group.

In addition, the stranger is not constrained by convention or precedent. Finally, Simmel described the tendency to perceive strangers in more abstract terms. With the stranger, one has only certain general qualities in common, whereas with more closely related group members our relationships are more detailed.

In the United States, the social scientists' concern with the phenomenon of the newcomer in the group may be traced at least to the time of the formal recognition of American sociology. The original justification for awarding academic status to sociology was the study of immigration, a problem with which no other social scientist dealt at that time (Shils, 1948). More recently, the massive population shifts particularly to Israel, Australia, and the United States, and within the United States have stimulated renewed interest in the problem of the newcomer and the host group under conditions of changing group composition (Eisenstadt, 1951, 1952; Richardson, 1957a; Taft, 1957; Weinberg, 1953). Within the United States alone, it has been estimated that one out of every five families moves each year (Rossi, 1955).

The early immigration studies relied perforce on historical documents, case studies, demographic data, and informal observations. The more recent studies tend to rely on survey research techniques. Other investigations have employed criteria of assimilation such as vocabulary changes (Richardson, 1957) and objective measures of status (Eisenstadt, 1951). Moreover, some of the programs have developed well-formulated theoretical frameworks from which, presumably, experimental programs will emanate (Taft, 1957).

Summarizing some of the results of these later immigration studies, it has been found that: (*a*) The relations between the primary groups and the new social and cultural systems are mediated by elites (Eisenstadt, 1951). (*b*) ". . . successful adjustment is associated with free (rather than forced) immigration, reasonable expectations of new life conditions, weak ties with former environment, and youth [Eisenstadt, 1956, p. 23]." (*c*) "[There is] a strong positive correlation between the general state of health and adjustment to the new country . . . [Weinberg, 1953, p. 63]." (*d*) Low family cohesiveness, greater similarity between host and donor communities, male rather than female, and relative youth were positively related to the kind of attitude possessed by the immigrants (Eisenstadt, 1952).

It is apparent that these findings concern, almost exclusively, the characteristics of the newcomer in a given culture or host group. Information is lacking which compares the assimilation process in groups with varying characteristics. This is one of the major limitations of immigration research with reference to the more general problem of the assimilation of the newcomer in the group. Necessarily this limitation of the immigration studies also restricts the scope of the theoretical structure based upon these studies.

The research discussed up to this point has concerned the microcosmic problems of immigration. However, more recently, research has become at once more microcosmic and more generic; the newcomer phenomenon has been explored with reference to smaller social units and wider variety of milieu. Studies were found concerning the assimilation of the foreign college student (Zajonc, 1952), the new member of the United States Supreme Court (Snyder, 1958), the new recruit in an Army platoon (Stouffer et al., 1949), and the new business leader (Christensen, 1953). Several comparative studies of animal behavior (King, 1954; Potter & Allee, 1950) and of laboratory experiments (Ziller & Behringer, 1960; Ziller, Behringer, & Goodchilds, 1962) were also found.

The microcosmic approach to problems of changing group composition enables the experimenter to employ small group techniques which have been developed rather extensively during the last fifteen years. Recognition of the advantages of small group research methods in this area have been remarkably tardy. Yet, there have been exceptions. In an early experiment by Lewin, Lippitt, and White (1950), the problems of leadership succession were explored but only as a peripheral interest.

In a more recent experiment with small groups (Ziller & Behringer, 1961), the assimilation process of the new child in the group was traced by means of sociometrics over an 11-week period in 28 elementary school classes. The results suggested that the newcomer's popularity during the 11-week period described a U-shaped curve with a marked decline in the early weeks. A similar phenomenon was observed with regard to Fulbright grantees visiting the United States (Lysgaard, 1955).

The results of a program of research concerning the acceptance of a new reference group has been reported recently by Hartley

(1958). In general, the results were derived from a response-response experimental design in which individuals in a new reference group (college or the Navy) completed questionnaires concerning individual characteristics on the one hand and acceptance of the new reference group on the other. The following hypotheses were supported.

Acceptance of a new group as a reference group was found to be associated with: (*a*) a tendency to perceive the members of that group as exhibiting similarity to each other and to the observer in tastes, goals, and objectives, and as offering each other mutual support; (*b*) a tendency to perceive the group or unique entity; (*c*) a conviction of clear awareness of the group's processes and functions; (*d*) the host group's relative position in the individual's hierarchy of groups; (*e*) the number of other reference groups; (*f*) positive expectations of a group; (*g*) the general tendency to associate with groups; (*h*) the relative length of anticipated contact with group; (*i*) positive rather than negative general social response patterns.

Newcomb (1961) conducted a longitudinal study of the acquaintance process and, particularly, mutual attraction among 17 male college newcomers and strangers who had agreed to live together for four months. Underlying the study was a system balance theory based on Heider's (1958) interpersonal framework. Thus it was proposed that a strain in an interpersonal system develops when two persons who are attracted to each other are not in agreement with regard to the attraction of a third person. One of the common processes of achieving balance in the system was to change attraction in the direction of those with whom one was in closer agreement with regard to attitudes and evaluations of others. Some of the conclusions reached by this series of studies include: (*a*) Agreement concerning relatively stabile attitudes during the acquaintance process becomes a significant determinant of attraction preferences. (*b*) Nonauthoritarians tend to become more accurate in estimating their fellows' orientations with increasing opportunity for obtaining information. (*c*) With increasing acquaintance, the general trend is toward the development of more high-attraction units of larger size (high-attraction pairs tend to build up into high-attraction triads, tetrads, and even larger units). (*d*) Interpersonal relationships of attraction stabilize at relatively early stages of development as individuals cease to acquire new information about each other.

Again, it is noted that the results of the preceding series of studies concern exclusively the characteristics of the newcomer to the group, since the characteristics of the host group are not varied under this type of design. Indeed, in Newcomb's (1961) study the newcomer's and the host group's characteristics were confounded. A more systematic and inclusive analysis of the newcomer phenomenon is demanded.

As a preliminary step in the analysis of the newcomer phenomenon, a framework of the salient components in the assimilation process may facilitate more comprehensive experimental inquiries. The components of the phenomenon appear to be: (*a*) the characteristics of the host group, (*b*) the characteristics of the donor group, (*c*) the characteristics of the newcomer, (*d*) the relationship between the host and donor group, (*e*) introduction and orientation procedures involving the newcomer, and (*f*) the bases for changes in group composition.

Host Group

In analyzing the various components of the assimilation process, we will have recourse to a decision-making model in which the rewards and cost relative to the newcomer and the host group are evaluated in juxtaposition. With reference to the personality of the host group, it is broadly hypothesized that sharply defined group boundaries tend to increase the cost of assimilation of new members both to the member and to the host group.

The hypothesis is almost trivial for its simplicity. The hypothesis simply holds that a broadly defined category can embrace a great many more objects than a narrowly defined category, and the cost to the group of establishing the criteriability of the object is diminished, thereby reducing the difficulty of assimilation both for the object and the perceiver (local group). A wide category host group, by necessity, places the onus of definition of role on the new member rather than on the system. Since selection science is rather inexact, this provides a degree of tolerance for the system which should facilitate integration. Moreover, by broadly defining the group, a set for tolerance rather than a set for criticism is created. The newcomer is not surveyed with a jaundiced eye but rather with a positive attitude toward idiosyncracies.

A wide-category group as defined here may be created in several ways. Group size is associated with the permeability of group boundaries and the open group in general (simply on a probability basis; e.g., the greater probability of a death), yet the perception and significance of group size is itself associated with a number of variables such as member homogeneity, group structure, and the personality dimensions of the members. The simplicity of group size as a concept is specious. Group size is a factor in consideration of the permeability of group boundaries, but the concept is of limited value in theory construction and experimentation.

In many respects, the open group derives from the wide-category concept. An open group is defined by the personnel which the group did have, does have, and will have and is necessarily broad in scope. Thus, it may be anticipated that new members are more readily assimilated in open groups. In a closed group the newcomer enters a far more complex, interrelated system of elements than in the open groups, and assimilation is accordingly accomplished with reduced facility (Ziller, Behringer, & Jansen, 1961).

Another way of creating group conditions conducive to ease of assimilation suggests itself. Recently, Pettigrew (1958) described the measurement of a "category width" variable; subjects tended to be consistently broad, medium, or narrow in their estimates of the ranges of the phenomena under consideration. Although the interpersonal-relations correlates of the instrument have not been explored to date, it may still be hypothesized that groups composed of members with wide category width indices as opposed to those with low indices will tend to assimilate a new member more readily. Similarly, a newcomer with wide category width tendencies may also be expected to accept membership in a new group more rapidly and thoroughly.

At a more general level, it may be proposed that highly differentiated or complex groups and newcomers facilitate assimilation due to the increased probability of matching a facet of the individual and the group.

Donor Group

The relationship of the characteristics of the donor group to those of the host group is a second component in the assimilation process. Again, quasi-statistical theory may be invoked to support the hypothesis that greater similarity between the host and donor groups tends to result in reduced assimilation difficulties. Again, if the host and donor groups are similar with regard to culture, goals, personnel, and structure, transition difficulties are avoided and the need for marked changes either by the newcomer or group is omitted.

This hypothesis is supported in part by the results reported earlier with reference to the immigration studies in Israel and Hartley's (1958) studies covering the assimilation of the new college student. Moreover, Wentholdt (1956) in studying the adjustment of Dutch immigrants to New Zealand suggests that successful personality readjustment of an immigrant to a "host" society varies with the degree of self-reorganization required by that society and the degree of preparedness on the part of the immigrant to have such reorganization take place.

Still, the shadow of the donor group tends to retard the growth of an integral relationship between the newcomer and the host group. In Christie's (1954) study of the transition from civilian to Army life, it was concluded that greater distance from home and limited contact with family and friends are more conducive to adjustment to military life during the basic training phases of the program. The expeditious use of isolation for facilitating the assimilation process of newcomers to the United States Coast Guard Academy is discussed by Dornbusch (1955).

Weiss (1963) used Miller and Dollard's (1941) social learning theory to analyze the relationship between reward and punishment associated with membership in one social movement and subsequent recruitment in other social movements. It was hypothesized that "the greater the punishment which led to defection, the less similar to the original movement is the new movement most likely to be chosen." Perhaps successive partners in courtship would be somewhat more amenable to experimentation than social movements.

The Newcomer

The last major component of the assimilation process is the newcomer. As indicated earlier, because of the ease of study, this component has been subjected to detailed analysis. Again, the studies by Eisenstadt (1952) and Hartley (1958) are cited. Without repeating the findings in detail, these studies suggest the existence of a cosmopolitan personality, a complex of behaviors which enables the individual to adjust to a variety of new social situations of changing social environments.

Merton's (1949) analysis of the "cosmopolitan" as opposed to the "local" is pertinent here. The cosmopolitans had lived in a variety of places and were newcomers to the town. The locals were usually natives and had seldom been away except for brief periods. The cosmopolitans were not members of as many groups as the locals. They stressed the importance of the qualities of their friends rather than the number of people they knew. Moreover, where the locals tended to join organizations for the purpose of extending their range of contacts, the cosmopolitans tended to join organizations in which special skills and knowledge were important.

It is quite conceivable, however, that the community behavior of these contrasting people stems from some underlying personality characteristics, again related to complexity of the self-concept. In terms of the aforementioned studies the cosmopolitan appears to be a xenophile; is not grossly nostalgic or sentimental; is liberal; in perception and concept formation characteristically employs wide category width and a large number of categories rather than dichotomous divisions; views the world positively; tends to be a risk taker; is not cynical; and is not prejudiced. The cosmopolitan in contrast to the local, is more highly transferable and his abilities are more highly transferable.

Transferability implies, as a very base, a positive attitude toward change, or at least, a lower resistance to change. In addition, since a variety of persons may be encountered who may be arranged along a normal curve on a number of dimensions, the probability of adjustment in new groups is probably increased if the mobile individual is located near the mean of the population on a number of these same dimensions. Thus, if the normal population tends to be only moderately prejudiced toward Negroes, for example, the highly prejudiced or highly pro-Negro newcomer will tend to have more difficulty adjusting to and gaining acceptance in a number of groups. This may suggest that individuals who successfully adjust to a variety of new situations tend to be conformists in relation to the behavior of the greatest number of persons in the population in which they are likely to move; just as the locally oriented individual is a conformist in relation to the behavior of the greatest number of persons in his immediate environment.

The discussion concerning the assimilation of the newcomer in the group has indicated that the change involved is more or less disturbing to either the group or the individual depending upon the characteristics of the group or the individual. It is now suggested that a change in group membership requires a reevaluation of the host group or the self. Under conditions of group stability less acceptable characteristics of the individual or of the group may be dissociated; the individual and the group through collusion may avoid confrontation of issues which are known to be disturbing. The introduction of a change, however, and a change of personnel in particular, forces a confrontation with the self. The newcomer mirrors the group to itself; and the group, in turn, mirrors the individual to himself. The confrontation is inherent in the initial assimilation process. Exposure must precede acceptance or adjustment in whatever form it takes. Thus, the individual or group experiencing difficulty in self-acceptance will have difficulty adjusting to changes in group membership; unresolved but compartmentalized or dissociated problems are reflected in interaction with the uninformed and unprepared other.

Continuous confrontation of the self either by the group or by the individual is presumed to generate countercathexis and an inward orientation which may detract from the general effectiveness of the organism to cope with day to day problems. Thus, a press is posited whereby both the stranger and the host group act to introject the other in order to avoid protracted mirroring or self-confrontation.

Leadership Changes

A special case of changes in group membership concerns changes in leadership or the group power structure. While there has been some speculation concerning the problem, surprisingly little research has been reported. Organizations are reluctant to permit analyses of group behavior during the critical transition period which accompanies leadership changes.

Gouldner (1954) reports a case study of leadership succession. He summarizes the problems of leadership succession as: (*a*) interaction between bearers of different values (the leader and the workers), (*b*) ambiguous canons of legitimacy, (*c*) unrequited expectations projected from the predecessor, (*d*) decline of informal interaction across status lines, (*e*) hiatus in the chain of command, (*f*) shortcircuited communications, (*g*) challenge to managerial legitimacy, (*h*) degeneration of motives for obedience.

Various acts are open to the new leader to establish organizational equilibrium and leader stability. Thus Gouldner observes that the new leader has recourse to "strategic replacement." A new informal social circle is formed about him by hiring new key personnel and through internal promotions.

Several field studies have been reported concerning leadership succession. Trow (1961) reported a 4%–5% succession rate for presidents of a random sample of small shoe, brewing, and automotive industries. In a study of succession among school superintendents, Carlson (1961) noted that promotions are internal when employers wish to maintain the system; but that outside promotions are designed for innovation.

Although the data and conclusion are uncomfortably disparate, Grusky (1961) suggests that bureaucracies require periodic succession at the top if they are to adapt adequately to their environment, since this is the only means of infusing a degree of flexibility which bureaucracies especially require. He also suggests, however, that bureaucracies not only require more, but also can stand more succession at the top levels because the position is well defined in terms of group structure, power, and communication channels.

Laboratory experiments have been relatively few. Pryer, Flint, and Bass (1962) found that successful groups changed leaders less often.

Consideration of leadership succession under various conditions of group structure and conditions of success and failure suggests that organizations in which responsibility for group action is shared by the members tend to perpetuate the attitudes and outcomes of the group following a change in formal leadership; whereas organizations in which power is centralized undergo severe changes following leadership succession. Thus, under conditions of group success followed by leadership change, democratic groups in contrast to autocratic groups have a greater probability of preserving the success pattern. Similarly, however, under conditions of failure, organizations with decentralized power may be expected to continue the failure pattern, whereas organizations with centralized power under conditions of leadership change are more likely to break with the past and improve the probability of success. Thus, it is proposed that democratic organizations are designed to perpetuate success, whereas autocratic organizations are designed to adjust to failure.

REFERENCES

Allport, F. H. *Social psychology.* Boston: Houghton Mifflin, 1924.

Anonymous. The making of the infantryman. *American Journal of Sociology*, 1946, **51**, 376–379.

Aronson, E., & Mills, J. The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, 1959, **59**, 177–181.

Becker, H. *Through values to social interpretation.* Durham: Duke Univer. Press, 1950.

Berlyne, D. E. *Conflict, arousal and curiosity.* New York: McGraw-Hill, 1960.

Bernard, J. *Remarriage: A study of marriages.* New York: Dryden Press, 1956.

Bertalanffy, L. V. The theory of open systems in physics and biology. *Science*, 1950, **3**, 23–29.

Brzezinski, Z., & Huntington, S. P. *Political power: USA/USSR.* New York: Viking Press, 1964.

Caplow, T., & McGee, R. J. *The academic market place.* New York: Basic Books, 1958.

Carlson, R. O. Succession and performance among school superintendents. *Administrative Science Quarterly,* 1961, **6**, 210–227.

Cartwright, D. Toward a social psychology of groups. Unpublished report, Research Center for Group Dynamics, Ann Arbor, Michigan, 1953.

Chapin, E. S. The optimum size of institutions: A theory of the large group. *American Journal of Sociology,* 1957, **62**, 449–461.

Chapple, E. D., & Coon, C. S. *Principles of anthropology.* New York: Holt, 1942.

Christensen, C. *Management succession in small and growing enterprises.* Andover, Mass.: Andover Press, 1953.

Christie, R. Transition from civilian to Army life. *Human Resources Research Office Technical Report,* 1954, No. 8.

Darley, J. G., Gross, N. U., & Martin, W. E. Studies in group behavior: The stability, change, and interrelations of psychometric and sociometric variables. *Journal of Abnormal and Social Psychology,* 1951, **46**, 565–576.

Dornbusch, S. M. The military academy as an assimilating institution. *Social Forces,* 1955, **33**, 316–332.

Dynes, D. R. Rurality, migration, and sectarianism. *Rural Sociology,* 1956, **21**, 25–28.

Eisenstadt, S. N. The place of elites and primary groups in the absorption of new immigrants in Israel. *American Journal of Sociology,* 1951, **57**, 222–231.

Eisenstadt, S. N. The process of absorption of our immigrants in Israel. *Human Relations,* 1952, **3**, 223–246.

Eisenstadt, S. N. Inter-generation tensions in Israel. *International Social Science Bulletin,* 1956, **8**, 1–23.

Festinger, L. *A theory of cognitive dissonance.* Stanford: Stanford Univer. Press, 1957.

Frenkel-Brunswik, E. Intolerance of ambiguity as an emotional and perceptual personality variable. *Journal of Personality,* 1949, **18**, 108–143.

Gerurd, M. W., & Dukette, R. Techniques for preventing separating trauma in child replacement. *American Journal of Orthopsychiatrics,* 1954, **24**, 111–127.

Gouldner, A. W. *Patterns of industrial bureaucracy.* Glencoe, Ill.: Free Press, 1954.

Gouldner, A. W. The norm of reciprocity: A preliminary statement. *American Sociological Review,* 1960, **25**, 161–178.

Grusky, O. Corporate size, bureaucratization, and managerial succession. *American Journal of Sociology,* 1961, **67**, 261–269.

Hartley, Ruth E. The acceptance of new reference groups: Final report, 1958, Contract Nonr-1597 (pl), Group Psychology Branch of the Office of Naval Research.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Hemphill, J. K. Situation factors in leadership. *Ohio State University Studies, Bureau of Educational Research Monographs,* 1950, No. 32.

Herbst, P. G. The analysis of social flow systems. *Human Relations,* 1954, **7**, 327–336.

Hollander, E. P. Conformity, status and idiosyncrasy credit. *Psychological Review,* 1958, **65**, 117–127.

King, J. A. Closed social groups among domestic dogs. *Proclamation of the American Philosophical Society,* 1954, **98**, 327–336.

Lewin, K. *Field theory in social science.* New York: Harpers, 1951.

Lewin, K., Lippitt, R., & White, R. K. Patterns of aggressive behavior in experimentally created social climates. *Journal of Social Psychology,* 1950, **3**, 155–174.

Lysgaard, S. Adjustment in a foreign society: Norwegian Fulbright grantees visiting the United States. *International Social Science Bulletin,* 1955, **7**, 45–51.

Machiavelli, N. *The prince and the discourses.* New York: Modern Library, 1940.

McVey, F. L., & Hughes, R. M. *Problems of college and university administration.* Ames: Iowa State College Press, 1952.

March, J. G., & Simon, H. A. *Organizations.* New York: Wiley, 1958.

Merton, R. Patterns of influence: A study of interpersonal influence and of communications behavior in a social community. In P. F. Lazarsfeld & Stanton (Eds.), *Communications research, 1948–1949.* New York: Harper, 1949.

Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

Mills, T. *Group structure and the newcomer.* Oslo, Norway: Oslo Univer. Press, 1957.

Newcomb, T. M. *Social psychology.* New York: Dryden Press, 1950.

Newcomb, T. M. *The acquaintance process.* New York: Holt, Rinehart, & Winston, 1961.

Pettigrew, T. F. The measurement and correlates of category width as a cognitive variable. *Journal of Personality,* 1958, **26**, 532–544.

Potter, H., & Allee, W. C. Some effects of experience with breeds of gallus on behavior of hen toward strange individuals. *Physiological Zoology,* 1950, **26**, 147–161.

Pryer, Margaret W., Flint, A. W., & Bass, B. M. Group effectiveness and consistency of leadership. *Sociometry,* 1962, **25**, 391–397.

Redfield, R. The folk society. *American Journal of Sociology,* 1947, **52**, 293–308.

Redfield, R. *The primitive world and its transformations.* New York: Cornell Univer. Press, 1953.

Richardson, A. The assimilation of British immigrants in Australia. *Human Relations,* 1957, **10**, 157–166. (a)

Richardson, A. Some psycho-social characteristics of satisfied and dissatisfied British immigrant skilled manual workers in Western Australia. *Human Relations,* 1957, **10**, 235–248. (b)

Rossi, P. H. *Why families move.* Glencoe, Ill.: Free Press, 1955.

Shepard, H. A. Research teams in industrial laboratories. Cambridge: Massachusetts Institute of Technology, School of Industrial Management, 1955. (Hecto)

Shils, E. A. *The present state of American sociology.* Glencoe, Ill.: Free Press, 1948.

Snyder, E. C. The supreme court as a small group. *Social Forces,* 1958, **36**, 232–238.

Sorokin, P. A. *Social mobility.* New York: Harper, 1927.

Stouffer, S. A., et al. *The American soldier: Combat and its aftermath.* Princeton: Princeton Univer. Press, 1949.

Taft, R. A psychological model for the study of social assimilation. *Human Relations,* 1957, **10**, 141–156.

Trow, D. B. Membership succession and team performance. *Human Relations,* 1960, **13**, 259–268.

Trow, D. B. Executive succession in small companies. *Administrative Science Quarterly,* 1961, **6**, 210–227.

Weinberg, A. A. Problems of adjustment of new immigrants to Israel. *Bulletin of the World Federation of Mental Health,* 1953, **5**, 57–63.

Weiss, R. F. Defection from social movements and subsequent recruitment to new movements. *Sociometry,* 1963, **26**, 1–20.

Wentholdt, R. *Personality readjustment among immigrants: A theoretical model.* (No. 7) New Zealand: Victoria University College Publications in Psychology, 1956.

Wolff, K. H. *The sociology of George Simmel.* Glencoe, Ill.: Free Press, 1950.

Wright, B. A. Altruism in children and the perceived conduct of others. *Journal of Abnormal and Social Psychology,* 1942, **37**, 218–233.

Zajonc, R. B. Aggressive attitudes of the stranger as a function of conformity pressures. *Human Relations,* 1952, **5**, 205–216.

Ziller, R. C. Scales of judgment: A determinant of the accuracy of group decisions. *Human Relations,* 1955, **8**, 153–164.

Ziller, R. C., & Behringer, R. D. Assimilation of the knowledgeable newcomer under conditions of success and failure. *Journal of Abnormal and Social Psychology,* 1960, **60**(2), 288–291.

Ziller, R. C., & Behringer, R. D. A longitudinal study of the assimilation of the new child in the group. *Human Relations,* 1961, **14**, 121–133.

Ziller, R. C., Behringer, R. D., & Goodchilds, J. D. Group creativity under conditions of success or failure and variations in group stability. *Journal of Applied Psychology,* 1962, **46**, 43–49.

Ziller, R. C., Behringer, R. D., & Jansen, Mathilda. The newcomer in open and closed groups. *Journal of Applied Psychology,* 1961, **45**, 55–58.

(Received December 8, 1964)

Psychological Bulletin
1965, Vol. 64, No. 3, 183–190

PSYCHOLOGICAL EFFECT OF EFFORT[1]

MICHAEL LEWIS

Fels Research Institute, Yellow Springs, Ohio

The present review explores the psychological effect of effort. While the present problem has received little attention, the studies of effort can be divided into two major areas: (a) Is effort an aversive stimulus, to be avoided in a choice situation and resulting in decrements in response strength? (b) Can effort once expended affect the stimulus-event (reward) associated with that effort? The available literature does not provide any conclusive statements concerning the aversive effect of effort. However, effort does seem to enhance the value of a stimulus associated with the expenditure of effort.

The problem of effort and its psychological effects has received little attention. This was due probably to the belief, stemming from Hull's (1943) theory that work was an aversive stimulus to be avoided given the choice, and that it resulted in decrements in response strength when applied. Recent investigation shows both assumptions to be in question and indicates that effort may even enhance the value of stimuli associated with it.

For the purpose of this review, the literature on effort or work will be divided into two main issues: (*a*) What are the effects of an organism's expenditure of effort upon the next effortful response? (*b*) Once an effortful response has been made and effort is no longer present, what are the effects of that effortful response?

Subsumed under (*a*) is the question: Does effort change the probability of making the next effortful response? This response could be a choice response or the continuation of a single response. To test whether effort changed the probability of choosing one or another response, investigators made one response (choice) more difficult than the other; for example, long versus short paths to a goal, jumping or not jumping over a barrier in order to reach a goal, and, finally, pushing light or heavy doors in order to reach a goal. To test whether effort changed the probability of the continuation of a response, investigators tested how long an organism would continue to respond (with no reinforcement) when the amount of effort it needed to expend was varied. These extinction types of experiments usually used Skinner boxes and differentially weighted levers.

In neither the choice nor the performance experiments were the experimenters interested in the effect of effort once effort *had been expended and was no longer expended*. Essentially, these experiments were interested in the question of whether the expenditure of effort was aversive. However, there were experimenters who were interested in the effect of effort *after* effort had been expended. These investigators were involved with the question of the effects of effort on the stimulus event (primary or secondary reward) associated with the expenditure of effort.

The following discussion will therefore be divided into two main areas: the effect of effort on a response, and the effect of effort, once expended, on the value of a stimulus event associated with the effortful response. The effect of effort on a response was measured in two ways: (*a*) preference (by choice) of the more difficult response over the easier one, and (*b*) the strength of the instrumental response. In studying the effect of effort on value of a stimulus event (primary or secondary reward), two measures of value were used: (*a*) choice; that is, choice of the stimulus event associated with high effort over low effort, and (*b*) the strength of the instrumental response in obtaining the stimulus event associated with high effort as opposed to low effort.

---

[1] This research was supported, in part, by Grant HD-00868 from the National Institute of Mental Health, United States Public Health Service.

EFFECT OF EFFORT ON A RESPONSE

Choice of Effortful Responses

The concept of effort was formally introduced by Hull as the psychological variable of work in his *Principles of Behavior* (1943). The notion of work or effort is to be found, however, in earlier experiments such as those by DeCamp (1920), Gengerelli (1931), Kuo (1922), and Tsai (1932). One of Hull's hypotheses concerning work was called the "least effort hypothesis" which stated:

> If two or more behavioral sequences, each involving a different amount of energy consumption of work, have been equally well reinforced an equal number of times, the organism will gradually learn to choose the less laborious behavior sequence leading to the attainment of the reinforcing state of affairs [Hull, 1943, p. 294].

In a later work (Hull, 1952), the hypothesis was stated as follows:

> Other things being equal, organisms which are presented with alternate paths in detouring about a barrier to an adient object will learn to prefer the one involving the shorter distance [p. 257].

Hull's view of effort has been open to two different kinds of tests. The first examines whether, if two responses are followed by the same reinforcement, the organism will tend to choose the less effortful of the two. For example, DeCamp (1920), Gengerelli (1931), and Kuo (1922) demonstrated that rats will learn to traverse the shorter of the two paths leading to the same reward. Waters (1937) varied the number of turns an animal had to make and showed that animals learned to choose the path with the fewest turns. Thompson (1944a, 1944b) used a symmetrical T maze with Skinner-type boxes at the ends of each arm. The levers required different degrees of effort to dispense the same reward. The animal learned to prefer the arm containing the lever easiest to press. Tsai (1932) demonstrated that rats preferred to push a light door rather than a heavier one in order to get food. He also found that animals preferred to climb over low rather than higher obstacles to obtain equivalent reinforcement. Finally, McCulloch (1934) tried to train rats to pull up the lighter of two trays containing food. His results were inconclusive in that the animals developed a position effect which made it impossible to ascertain the effects of the effort conditions. In a recent study by Chung (1965) the same problem was again studied. Pigeons were trained to depress two levers by pecking, each requiring 25 grams of force. After the training procedure the force required to depress one of the levers was changed so that one key was easier to depress than the other. Chung found that the rate of response on one lever was independent of the required force on the other lever. In other words, effortfulness was not an aversive stimulus, and therefore the animals showed no preference for the least effortful response. Thus, unlike the other experiments on choice, the animals showed no preference for the less effortful response. With the exception of this result, the data on choice appear to confirm Hull's viewpoint.

Strength of Instrumental Responses

Hull (1943) also viewed work in terms of its subtractive effect on performance and utilized it in several of his laws, particularly in relation to reactive and conditioned inhibition. He stated that "the net amount of functioning inhibitory potential resulting from a sequence of reaction evocations is a positively accelerated function of the amount of work involved in the performance of the response in question [p. 279]." Hull further believed that this inhibitory potential acted as a "primary negative motivation in that it had the innate capacity to produce a cessation of the activity which caused the state [p. 279]." One of the differences between reactive and conditioned inhibitory potentials was the duration of the negative motivational state. Extinction data were used, therefore, as a measure of the effect of effort on the effective reactive potential $({}_{S}E_{R})$ of a response. Hull believed that inhibitory potential $({}_{S}I_{R})$ was a function of work or effort, that is, the greater the work required for a response, the greater the inhibitory potential of that same response. Since inhibitory potential subtracts from the total effective reaction potential, those responses involving greater effort or work would presumably extinguish more quickly than those responses requiring less effort. Crutchfield (1939) trained rats to pull differentially weighted food cups in order to obtain food and found results that did not agree with Hull's expectation. During extinction he found generally that those animals pulling the heavier weights tended to extinguish more slowly than those animals pulling lighter weights.

Mowrer and Jones (1943) attempted to investigate the relationship between effort and resistance to extinction by training a group of rats to press down on a lever in order to receive a food pellet. The training procedure first conditioned all the animals to press down with 5 grams of pressure and then gradually increased the pressure required to 80 grams. The animals were then divided into three groups based on the amount of effort needed to depress the bar during extinction: 5, 42.5, or 80 grams. The results indicated a perfect negative relationship between effort during extinction and resistance to extinction and thus appeared to support Hull's hypothesis. But the experiment was difficult to interpret since the animals received 180 reinforced trials on the light bar during training, as compared to only 80 reinforced trials on the heavy bar. The results could therefore be explained on the basis of differential frequency of reinforcement during acquisition rather than on the basis of effortfulness of response during extinction.

Applezweig (1951) controlled for frequency of reinforcement during acquisition. He used five groups, each trained with a different bar-pressing weight: 10, 20, 30, 40, and 50 grams. Each acquisition group contained five animals and was divided so that one animal was extinguished on each of the five different bar-pressing weights. The results generally agreed with those of Mowrer and Jones—the greater the amount of effort needed to press the bar during extinction, the more quickly did the response extinguish, even among subjects who experienced a single common level of effort during acquisition. The results also indicated, however, that animals both trained and extinguished on the more effortful bar took longer to extinguish than those trained and extinguished on the least effortful bar. In order to explain this last finding, Applezweig argued that the feedback stimulation from the bar-pressing response was analogous to Hull's V construct (stimulus intensity dynamism), and that it was greater under conditions of high rather than low effort. He further stated that response strength should be greater in the high-effort group than in the low-effort group because of V's multiplicative relationship to other variables: ${}_sE_R = f\ ({}_sH_R \times D \times V)$.

Maatsch, Adelman, and Denny (1954), using a larger number of subjects under the same experimental conditions as those of Mowrer and Jones, tested for the same effect. Using the time required to obtain the last 10 rewards of the training period as a measure of habit strength, the authors found no difference among the three groups. The results of their extinction data also indicated no significant differences among the groups.

Capehart, Viney, and Hulicka (1958) also controlled for frequency of reinforcement during acquisition. These authors used 5-, 40-, and 70-gram bar weights and found that the number of responses emitted during extinction varied inversely with the effortfulness of the response. Quartermain (1965) repeated the Capehart et al. experiment but was unable to find any difference in the number of responses during extinction as a function of the effort required in depressing the bar.

Aiken (1957) had rats push open differentially weighted doors in order to get food and found no significant difference between groups during extinction. Solomon (1948) found that animals extinguished quicker when forced to jump 16 rather than 8 inches from one stand to another in order to receive food. It could be argued, however, that competing responses induced by the longer jump rather than differential effort were responsible for the differences in extinction between the groups. A jump of 16 inches might be more frightening than a jump of 8 inches, thus creating differential anxiety. In that the 16-inch jump associated with greater anxiety, one would expect more rapid extinction.

Finally, the studies of Lawrence and Festinger (1962) are relevant to the discussion of effort and rate of extinction. The authors found that when the slope of a straight alley maze was varied, animals who were required to run up a steep 45-degree slope (to exert greater effort) extinguished more slowly than animals who were required to run along a level runway. In another experiment, Lawrence and Festinger demonstrated that increased effort during partial reward intensified the effect of unrewarded experiences on resistance to extinction. The authors showed that effort did not produce faster extinction as would be predicted by Hull's formulations. Thus, because of inconsistent results, it cannot be concluded that effort decreases the strength of an instumental response.

EFFECT OF EFFORT ON VALUE

The previous experiments had little bearing on the value of the goal or of the reinforcer itself. They related primarily to the choice or strength of instrumental responses or paths to a goal. The experiments to be reported in this section are interested in the consequences of effort on a stimulus event once effort is expended. That is, does the expenditure of effort enhance a stimulus event *associated* with the effortful experience? The value of the stimulus event, either a primary or secondary reward, was tested either by choice or by the strength of the instrumental response necessary to obtain the reward. In any test of the value of the reward, the original effort conditions were not present.

Wright (1937) argued that the presence of a barrier between a subject and a desired goal increased the positive valence of that goal. In a series of experiments with children, Wright reported that their preference for an object was greater when the object was less accessible than when it was easy to reach. He had children indicate their preference both by their action (choice) as well as by their verbal report and used these as measures of the valence of the object. His experiment may be interpreted as indicating that *anticipated* effort enhances the preference or incentive value of a goal. It does not indicate, however, the effect of actually making the effortful response. Child and Adelshiem (1944) repeated Wright's experiment and reported contradictory findings. Their subjects preferred the more easily attainable goals. In a subsequent study, Child (1946) attempted to demonstrate that the preference for a goal more difficult to attain was the result of several factors, the most important of which was previous learning or socialization with respect to differential reinforcement for differential effort, such that the child came to value the stimulus properties of effort itself. He introduced variables such as age and sex and showed that the preference for the distant goal increased with age, and that boys were more likely to prefer it than girls. He attributed these phenomena to differential degrees of learning of a value structure in which effortful goals were sometimes preferred because of previous reinforcing consequences, rather than to a direct effect of effort upon valence.

Festinger's (1957, 1961) theory of cognitive dissonance is also relevant to the problem of effort and value. Festinger defined cognitive dissonance as occurring when two cognitive elements are not in agreement. When dissonance occurs, the organism is motivated to reduce the dissonance by changing one or both elements until they agree. In a paper by Festinger and Aronson (1960), the authors specifically discussed the theory's relevance to effort and value. They contended that when an organism obtains an "insufficient reward" relative to a particular expenditure of energy, there will be a tendency to discontinue the effortful activity or to attribute additional value to the activity or to its goal consequences. This development of an "extra preference" for the activity or its goal consequences is conceived as resulting from an internal process of dissonance resolution; that is, the receipt of a reward incommensurate with the effort expended in its attainment leads to a state of cognitive dissonance that the organism is motivated to reduce. In a study by Aronson (1961), preference was directly assessed following a series of unrewarded trials on tasks which required "considerable expenditure of energy" for one group and were "almost effortless" for another group. The high-effort group showed a clear preference over that shown by the low-effort group for irrelevant aspects (secondary reinforcement) of the goal situation.

Olds (1953) used children in an experiment attempting to show that "wanting-practice" affects the value of a secondary reinforcer. Practice in "wanting" was varied by using a task in which differences in expended effort could be introduced. This was done by varying the number of turns of a crank necessary to attain a reinforcer. In the habituation phase of the experiment, Olds trained the children to turn a crank three times in order to receive a chip. They did this for 20 days, receiving 8 chips a day. At the end of this period the subjects were divided into four groups. Group A now had to turn the crank 12 times to get a chip; Group B, 7 times; Group C, 3 times; and Group D, just once. Using the pressure exerted by the children in turning the crank as the measure of the secondary reward value of the chips, Olds found that Groups A and B valued the chips more than Groups C and D. While the results may be interpreted as indicating that effort directly affects the value of a reinforcer, they may also be interpreted as indicating that the chip value is constant, but that the children in Groups A and B turned with greater pressure simply because they experienced an increment of drive attributable to frustration.

Three recent and related experiments have been directly concerned with the problem of whether a stimulus event undergoes changes in value as a function of the effort expended during the response it reinforced. In one of these (Lewis, 1964a), first- and sixth-grade boys were subjected to either a high- or low-effort procedure in order to obtain 20 poker chips. Effort was varied by altering the number of turns of a crank handle necessary to obtain a single chip. Two kinds of indices of value were used: (*a*) preference value was assessed by determining the number of alternative chips (differing in color from the training chip) the child would exchange for the training chips; and (*b*) reinforcement value—the effect of presentation of a training chip on the probability of the occurrence of a subsequent new response—was assessed in a binary choice problem where the chips were used as the reinforcers for a correct response. While the preference value failed to show the effects of effort, the reinforcement value of the chips was affected so that those associated with high effort were more effective in increasing the probability of the occurrence of the desired response than were those associated with low effort.

In another study using animals as subjects, Lewis (1964b) again tested whether the reward value of an object is functionally related to the amount of effort required to obtain that object. Three experiments were conducted where rats were required to pull weights in order to obtain a minimal amount of a relatively unique food reward. Four weight groups were used: 5, 30, 55, and 80 grams. The weight that the subject pulled was constant from trial to trial. After the training periods the weights were removed and not used during the testing procedure. During the test phase of the experiments, value was assessed in three ways: the running speed to the food in the straight alley maze, the speed with which the animal consumed the food, and the amount of the food consumed in a free-feeding situation when the animal was satiated on another food. The results of all three experiments supported the hypothesis that animals expending the greatest effort in a training series develop the strongest preference for the stimulus event (food).

In a final experiment, Lewis (1965) tested the hypothesis that a secondary reinforcer associated with high effort would be preferred to the same secondary reinforcer associated with low effort. Preference was measured in two ways: (*a*) subject's choice in a T maze, and (*b*) time spent in the presence of the secondary reinforcer. Each animal experienced both high- and low-effort conditions associated with either a vertically or a horizontally striped area. Half of the subjects experienced high effort with vertical and low effort with horizontal stripes. The other half experienced high effort with horizontal and low effort with vertical stripes. During the test phase, the animal, without weights, was given a choice between the striped areas and the total time spent in each arm was recorded. There was no reinforcement given during the test trials. The results demonstrated that effort is effective in influencing the value of a secondary reinforcer. Thus, the effect of effort on value has been clearly seen in three experiments, although its effect was mild. Festinger (1961), in discussing the effect of effort, also reached the conclusion that for the rat, the effort variable has a weak influence on the shift in the value of a stimulus event.

THEORETICAL ISSUES

Existing within the body of psychological literature are several theoretical positions which can account for the fact that effort can affect the value of a stimulus event used as a reward for the effortful response, as well as explain the nondecremental effect of effort on a response. Four such positions are given below.

Past Experience

The history of the organism, either animal or human, may be such that responses involving differential effort have been associated with differential reinforcement; that is, greater effort may have been followed by greater reinforcement. This relationship between effort and reward may generalize to new situations where effort may induce expectations of high reward. For humans, and especially Western man, this has particular relevance. Max Weber, in his classic treatise *The Protestant Ethic and the Spirit of Capitalism* (1930), and others following him (e.g., McClelland, in *The Achieving Society,* 1961) have pointed out that work has acquired a special meaning in Western society and that, to some extent, it is valuable in proportion to effort expended and not merely to the consequences of the work. From a less behavioral point of view, one might also may that the "Protestant Ethic" is reflected in cognitions such as, "I worked harder for it; therefore, it is more valuable." This point is explicit in Festinger's (1957) theory of cognitive dissonance and work, which states that hard work for a small reward may enhance the value of the reward because one expects large rewards for hard work. The age and sex variables discussed by Child (1946) can reflect the fact that such cognitions are acquired through socialization.

Proprioceptive Cues

It could be argued that proprioceptive stimuli associated with high effort are more intense or perceptually articulate (i.e., have greater "S_D" value), and that their discriminative properties enhance their derived (secondary) value for any given amount of original external reinforcement. After training with differential effort, in any subsequent test of the value of a reinforcer in which a response at all similar to the training response is used, the apparent value of the external reinforcer might seem greater for a group trained under high effort. But this effect could be attributed to a generalization effect, with the high-effort conditions producing a greater generalization of response than the low-effort condition. Thus, the superior performance may not reflect an increased preference, but rather a differential generalization of response patterns based on differential proprioceptive feedback. Such a problem occurred when Lewis (1964b) used a test response similar to the training response. By using choice as a test response, a response completely different from the training response, Lewis (1965) eliminated this interpretive problem.

Frustration

A third problem in interpreting experiments on the effect of effort revolves around the concept of frustration. High effort might be viewed as interfering with a goal-directed response, resulting in frustration and increasing the general drive level of the organism (Brown & Farber, 1951; Hull, 1952; Sears, Maccoby, & Levin, 1957). This interpretation of effort in terms of frustration maintains that the effort itself did not directly affect the value of the external reinforcer, rather that it affected drive level and the consequent reinforcement value of drive reduction.

The same kind of interpretative problem exists where frustration is defined as withdrawal or delay of expected reward (Amsel, 1958; Holder, Marx, Holder, & Collier 1957; Wagner, 1959). Old's (1953) work is particularly open to this kind of interpretation in that the failure to receive the accustomed reward could have led to frustration and therefore to increased drive.

Competing Responses

The last problem in interpreting the effect of effort is the attentional factor. Lewis (1964a, 1964b, 1965) found that effort tended to hamper any movement or activity not directed toward obtaining the goal. In the animal experiments, high-effort animals moved directly toward the goal area, while low-effort animals jumped about and explored the straight-alley maze. The high-effort as compared to the low-effort children also seemed to be more involved in obtaining the chips, that is, in looking at the dispenser and in turning the crank, rather than looking around the experimental room. One might view effort, then, as inhibiting those responses that are not compatible with obtaining the stimulus event, thereby increasing the probability of the occurrence of the goal-directed response. While this explanation can best account for effort differences in a test situation having a response similar to the training response (see Lewis, 1964b), it might also be applicable to other situations.

CONCLUSION

The preceding discussion makes it clear that the available literature does not provide any conclusive statements concerning the effects of effort. However, the data do lend general support to Hull's (1943) view that an organism will choose the least effortful of two responses. The evidence for Hull's belief that effort has a subtractive effect on performance is inconclusive. More data, in a variety of experimental settings, is needed before this issue is settled.

Finally, the data on the effect of the expenditure of effort on the value of a stimulus event associated with effort indicates that value can be enhanced by effort. The reasons for this effect are not clear, necessitating further investigation.

REFERENCES

Aiken, E. G. The effort variable in the acquisition, extinction and spontaneous recovery of an instrumental response. *Journal of Experimental Psychology*, 1957, **53**, 47–51.

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

Applezweig, M. H. Response potentials as a function of effort. *Journal of Comparative and Physiological Psychology*, 1951, **44**, 225–235.

Aronson, E. The effect of effort on the attractiveness of rewarded and unrewarded stimuli. *Journal of Abnormal and Social Psychology*, 1961, **63**, 375–380.

Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychological Bulletin*, 1951, **48**, 465–495.

Capehart, J., Viney, W., & Hulicka, I. M. The effect of effort upon extinction. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 505–507.

Child, I. Children's preference for goals easy or difficult to obtain. *Psychological Monographs*, 1946, **60**(4, Whole No. 280).

Child, I., & Adelsheim, E. The motivational value of barriers for young children. *Journal of Genetic Psychology*, 1944, **65**, 97–111.

Crutchfield, R. S. The determiners of energy expenditure in string-pulling by the rat. *Journal of Psychology*, 1939, **7**, 163–178.

Chung, Shin-Ho. Effects of effort on response rate. *Journal of the Experimental Analysis of Behavior*, 1965, **8**, 1–8.

DeCamp, J. E. Relative distance as a factor in the white rats' selection of a path. *Psychobiology*, 1920, **2**, 245–253.

Festinger, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

Festinger, L. The psychological effects of insufficient rewards. *American Psychologist*, 1961, **16**, 1–11.

Festinger, L., & Aronson, E. The arousal and reduction of dissonance in social contexts. In D. Cartwright & A. Zander (Eds.), *Group dynamics.* Evanston, Ill.: Row, Peterson, 1960.

Gengerelli, J. A. The principle of maxima and minima in animal learning. *Journal of Comparative and Physiological Psychology*, 1930, **11**, 193–236.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Holder, W. B., Marx, M. H., Holder, E. E., & Collier, G. Response strength as a function of delay of reward in a runway. *Journal of Experimental Psychology*, 1957, **53**, 316–323.

Kuo, Z. Y. The nature of unsuccessful acts and their order of elimination in animal learning. *Journal of Comparative and Physiological Psychology*, 1922, **2**, 1–27.

Lawrence, D. H., & Festinger, L. *Deterrents and reinforcement: The psychology of insufficient reward.* Stanford, Calif.: Stanford Univer. Press, 1962.

Lewis, M. The effect of effort on value: An exploratory study of children. *Child Development*, 1964, **35**, 1337–1342. (a)

Lewis, M. Some nondecremental effects of effort. *Journal of Comparative and Physiological Psychology*, 1964, **55**, 367–372. (b)

Lewis, M. The effect of effort on choice: The value of a secondary reinforcer. *Psychological Reports*, 1965, **16**, 557–560.

Maatsch, J. L., Adelman, H. M., & Denny, M. R. Effort and resistance to extinction of the bar-pressing response. *Journal of Comparative and Physiological Psychology*, 1954, **47**, 47–51.

McClelland, D. C. *The achieving society.* Princeton, N.J.: Van Nostrand, 1961.

McCulloch, T. L. Performance preferentials of the rat in force-resisting and spatial dimensions. *Journal of Comparative and Physiological Psychology*, 1934, **18**, 85–111.

Mowrer, O. H., & Jones, H. M. Extinction and behavior variability as functions of effortfulness of task. *Journal of Experimental Psychology*, 1943, **33**, 369–386.

Olds, J. Influence of practice on the strength of secondary approach drives. *Journal of Experimental Psychology*, 1953, **46**, 232–236.

Quartermain, D. Effect of effort on resistance to extinction of the bar-pressing response. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 63–64.

Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child-rearing*. Evanston, Ill.: Row, Peterson, 1957.

Solomon, R. Effort and extinction rate: A confirmation. *Journal of Comparative and Physiological Psychology*, 1948, **41**, 93–101.

Thompson, M. E. An experimental investigation of the gradient of reinforcement in maze learning. *Journal of Experimental Psychology*, 1944, **34**, 506–515. (a)

Thompson, M. E. Learning as a function of the absolute and relative amounts of work. *Journal of Experimental Psychology*, 1944, **34**, 506–515. (b)

Tsai, L. S. The laws of minimum effort and maximum satisfaction in animal behavior. *Monographs, National Research Institute of Psychology*, (Peiping, China), 1932, No. 1. (*Psychological Abstracts*, **6**:4329)

Wagner, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *Journal of Experimental Psychology*, 1959, **57**, 130–136.

Waters, R. H. The principle of least effort in learning. *Journal of General Psychology*, 1937, **16**, 3–20.

Weber, M. *The protestant ethic and the spirit of capitalism*. (Trans. by Talcott Parsons) London: Allen & Univin, 1930.

Wright, H. F. The influence of barriers upon strength of motivation. *Contributions to Psychological Theory*, 1937, **1**(Whole No. 3).

(Received February 5, 1965)

Psychological Bulletin
1965, Vol. 64, No. 3, 191–205

THE ROLE OF THE FAMILY IN THE DEVELOPMENT OF PSYCHOPATHOLOGY

GEORGE H. FRANK
University of Miami

Psychologists generally make the assumption that the experiences to which the individual is exposed over a period of time lead to the development of learned patterns of behavior. From this, psychologists have reasoned that the experiences the individual has in his early life at home, with his family, in general, and his mother, in particular, are major determinants in the learning of the constellation of behaviors subsumed under the rubric, personality, and in particular, the development of psychopathology. A review of the research of the past 40 yr. failed to support this assumption. No factors were found in the parent-child interaction of schizophrenics, neurotics, or those with behavior disorders which could be identified as unique to them or which could distinguish one group from the other, or any of the groups from the families of the controls.

As psychopathology came to be viewed as the consequence of the emotional experiences to which the individual was exposed, interest was focused on the earliest of such experiences, those that occur in the family. The human infant is born incapable of sustaining its own life for a considerable length of time following birth, and is, in consequence, dependent upon the mother or a mother substitute for its very existence. There is no wonder, therefore, that the mother-child relationship is a close one and is expected to be influential with regard to the psychological development of the child. Some explanations for the development of psychopathology have therefore focused on this particular relationship as the major etiological factor. Levy (1931, 1932, 1937, 1943) has described a pattern centering around "maternal overprotection," involving a constellation of attitudes which he felt contributed to the development of neurotic disorders, and Despert (1938) focused on a kind of mother-child relationship which seemed to her to be closely associated with the development of schizophrenia, a pattern which has come to be termed the "schizophrenogenic mother."

The hypothesis that the emotional climate of the interpersonal relationships within the family—and between the child and its mother in particular—has a decisive part in the development of the personality of the child would seem to have face validity. In part, support for this hypothesis may be gleaned from the data demonstrating the devastating effects of being brought up in the extreme interpersonal isolation that comes from *not* having a family (Beres & Obers, 1950; Brodbeck & Irwin, 1946; Goldfarb, 1943a, 1943b, 1943c, 1945a, 1945b; Lowrey, 1940; Spitz, 1945) or extreme social isolation within a family (Bartmeier, 1952; Davis, 1940). Moreover, it has been demonstrated that various specific emotional behaviors of the child seem to be correlated causally with factors in the home. For example, children who could be described as emotionally immature, who are dependent, fearful, negativistic, emotionally labile, etc., have had mothers described as worriers (Pearson, 1931), overattentive (Hattwick, 1936; Hattwick & Stowell, 1936), or punitive (McCord, McCord, & Howard, 1961; Sears, Whiting, Nowlis, & Sears, 1953; Watson, 1934). Children who were described as being overly aggressive were described as having come from homes where mothers were seen as overcontrolling (Bishop, 1951) or punitive (McCord et al., 1961; Sears, 1961).

The evidence thus far suggests that there is, in fact, a correlation between events in the parent-child relationship and resultant personality *traits*. The question arises as to whether there is evidence which supports the hypothesis that there is a correlation between events in the parent-child relationship and the resultant complex patterns of behavior which have been termed personality. More specifically, in light of the theories which relate personality development to social, (i.e., interpersonal) learning, the question is raised as to whether there is any consistent relationship between the emotional experience the child may have in the home and the development of personality pathology, that is, schizophrenia, neurosis, and behavior disorders. Towards this end, the findings of the research that has explored the psychological characteristics of the parents of these people will be analyzed in order to isolate those consistent characteristics of the parents that may emerge from study to study. The analysis will be done with regard to each major type of psychopathology as a group. Moreover, because psychological test data might yield different information than case history analysis, or direct observation of familial interaction as compared to attitudes as elicited by questionnaire, an attempt will be made to analyze the information gleaned from the studies in terms of the method of data collection within the specific psychopathological groupings.

SCHIZOPHRENIA

Case History

One of the classical methods of data collection in the study of psychiatric illness is the case history, the information for which has generally been gathered by other professionals. The individual conducting a piece of research notes the material in the folders and draws conclusions from the collation of these observations.

In so doing, Despert (1938) observed that approximately 50% of the mothers of a sample of schizophrenic children, generally between the ages of 7 and 13, had been described as aggressive, overanxious, and oversolicitous and were considered to be the dominant parent. Clardy (1951) noted that 50% of the 30 cases of children between the ages of 3 and 12 diagnosed as schizophrenic had families characterized as overprotective and yet basically rejecting. Frazee (1953) noted the presence of this constellation particularly when the families of schizophrenics were compared with the families of children diagnosed as behavior disorders. Canavan and Clark (1923) and Lampron (1933) noted that 30% of the children of psychotics were themselves emotionally disturbed. Huschka (1941) and Lidz and Lidz (1949) noted that over 40% of their sample of schizophrenics had parents who were psychotic or neurotic. Bender (1936, 1937) and Frazee (1953) noted the high incidence of psychopathology in the children of psychotic parents. Preston and Antin (1932), on the other hand, found no significant differences in the incidence of psychosis and neurosis as a function of parents who were psychotic as compared to parents who were "normal," and Fanning, Lehr, Sherwin, and Wilson, (1938) found that 43% of the children of mothers who were psychotic were observed to be making an adequate social and personal adjustment, with only 11% of that sample classified as maladjusted.

Lidz and Lidz (1949) found that 40% of their sample of schizophrenic patients were deprived of one parent by divorce or separation before they were 19. Plank (1953) found that 63% of his sample of schizophrenics had families where one parent was absent either due to death or marital separation. Wahl (1954, 1956) found that there was a greater incidence of parental loss and rejection early in life for schizophrenics as compared to normals, and Barry (1936) found that from the case histories of 30 rulers adjudged, post facto, insane, 80% of them had lost one of their parents by the time they were 18. However, Barry and Bousfield (1937) found that the incidence of orphanhood in a psychiatric population (19 out of 26) was not much different from the incidence of orphanhood in a normal population (19 out of 24). Moreover, Oltman, McGarry, and Friedman (1952) found that the incidence of broken homes and parental deprivation in the families of schizophrenics (34%) was not very different from that found in the families of hospital employees (32%), alcoholics (31%), and manic-depressives (34%); indeed, in their sampling, neurotics (49%) and psychopaths (48%) showed a greater incidence. Other studies have found that the incidence of broken homes in the history of neurotics is between

20% (Brown & Moore, 1944) and 30% (Madow & Hardy, 1947; Wallace, 1935), and Gerard and Siegel (1950) found no particular incidence of broken homes in the family history of their sample of schizophrenics.

Psychiatric Interview

Another classical method of obtaining information regarding the individual with whom patients have been living is by having interviews with them directly. The quality of the mother-child relationship is then inferred from what the interviewee says. From this research, an overwhelming number of studies (Despert, 1951; Gerard & Siegel, 1950; Guertin, 1961; Hajdu-Gimes, 1940; Kasanin, Knight, & Sage, 1934; Lidz, Cornelison, Fleck, & Terry, 1957a, 1957b, 1957c; Lidz, Cornelison, Terry, & Fleck, 1958; Lidz & Lidz, 1949; Lidz, Parker, & Cornelison, 1956; Tietze, 1949; Walters, 1938) describe a familial pattern characterized by a dominant, overprotective, but basically rejecting mother and a passive, ineffectual father. Yet the data in the study by Schofield and Balian (1959) reflected similarity rather than differences in the families of schizophrenic and nonpsychiatric (general medical) patients, and the data of Gerard and Siegel (1950) indicated that the schizophrenics in their study, according to interpretation of the data gleaned from the interviews, received adequate breast feeding, had no history of particularly difficult toilet training or of obvious feeding problems, did not come from broken homes, and apparently were not unduly rejected or punished. Another factor which seems to emerge from the studies is that a dominant characteristic of the family life of schizophrenics is a quality of inappropriateness of thinking and behaving which seems to infiltrate the entire atmosphere (Fleck, Lidz, & Cornelison, 1963; Lidz et al., 1957b, 1957c; Stringer, 1962). Meyers and Goldfarb (1962), however, found that only 28% of the mothers of 45 children diagnosed as schizophrenic and only 12% of the fathers were themselves manifestly schizophrenic.

Psychological Evaluation

Attitude Questionnaires. One of the most widely used questionnaires in this area of research has been the Shoben (1949) Parent-Child Attitude Survey. The Shoben scale consists of 148 items which measure the dimensions of parental rejection, possessiveness, and domination. From the administration of this attitude survey, Mark (1953) and Freeman and Grayson (1955) reported significant differences in attitudes toward child rearing between mothers of schizophrenics and mothers of normal children. In comparison with the mothers of the control subjects, the mothers of schizophrenic patients (Mark, 1953) were revealed as inconsistent in their methods of control. They described themselves as being, at times, overrestrictive and controlling of behavior, but in some instances lax. They frowned on sex play and tended to keep information regarding sex from their children; they also seemed to frown on friends for their children. Their relationship to their children appeared inconsistent; they described what could be interpreted as excessive devotion and interest in the child's activities while at the same time revealing a notable degree of "cool detachment." Freeman and Grayson (1955) found that in comparison to mothers of students in an undergraduate course, mothers of 50 hospitalized schizophrenics (ages 20 to 35) tended to reveal themselves to be somewhat more possessive, but inherently rejecting of their children, and particularly disturbed about sexual behavior in their children. However, according to these same data, the mothers of schizophrenic patients did not reveal themselves to be more dominant, dogmatic, or inconsistent in their attitude than the controls. But most important was the fact that item analysis of these data revealed that the attitudes of the mothers of the schizophrenics and of the controls were distinguished on only 14 of the items, and then, in general, there was so much overlap that even on these items the statistical significance was contributed by a small percentage of each group. Freeman, Simmons, and Bergen (1959) included four items from the Shoben scale among a larger sample of questions posed to parents. These items had been derived from a previous study (Freeman & Simmons, 1958) and were included in the second study because they were the only ones in the first study which were found to discriminate between the attitudes of mothers of schizophrenic patients and those of mothers of normals. The items are:

1. Parents should sacrifice everything for their children.
2. A child should feel a deep sense of obligation always to act in accord with the wishes of his parents.
3. Children who are gentlemanly or ladylike are preferable to those who are tomboys or "regular guys."
4. It is better for children to play at home than to visit other children.

Freeman et al. (1959) found no capacity for these items to differentiate the attitudes of the mothers of schizophrenics from those of other individuals with severe functional disorders.

Zuckerman, Oltean, and Monashkin (1958) utilized another attitude scale, the Parental Attitude Research Inventory (PARI, developed by Schaefer and Bell in their work at the Psychology Laboratory at NIMH). The PARI was administered to mothers of schizophrenics, and it was found that only one item distinguished between their attitudes and those of mothers of normal children. The mothers of schizophrenics tended to describe themselves as being stricter than did the mothers of nonschizophrenic children.

The minimal discrimination value of the several attitude scales should be noted. This would seem to reflect either minimal capacity of the scales to make such distinctions or little in the way of measurable differences between the groups. In either case, it is very difficult to evaluate the meaning of these data since the attitudes of the mothers of schizophrenics seemed to be distinguished from the attitudes of the mothers of neurotics on only a few items (the obtained number of differences did not even exceed that expected by chance alone).

Projective Tests. Several studies presented Rorschach data on the mothers of schizophrenic patients (Baxter, Becker, & Hooks, 1963; Prout & White, 1950; Winder & Kantor, 1958). In comparison to those of the mothers of normals, the Rorschach protocols of the mothers of the schizophrenic patients were undistinguished as regards the general degree of immaturity (Winder & Kantor, 1958) and the use of defences which are essentially reality distorting, namely, denial and projection (Baxter et al., 1963).[1] However, Prout and White did find more pure color without form and less human and animal movement and shading responses in the Rorschach protocols of mothers of schizophrenic boys as compared to the mothers of a comparable group of boys randomly selected from the community. Perr (1958) found that the parents of schizophrenic children gave responses to the Thematic Apperception Test (TAT) little distinguished from those of parents of normal children, and Fisher, Boyd, Walker, and Sheer (1959) found that the TAT and Rorschach protocols of the parents of schizophrenic patients were measurably different from those of the parents of nonpsychiatric (general medical) patients, but they were not distinguishable from the protocols of the parents of neurotic patients. The mothers of the schizophrenics revealed a higher degree of perceptual rigidity, greater incidence of indicators of maladjustment on the Rorschach, and less definitely conceived parental images on the TAT than the mothers of the normals.

Direct Observation of Interpersonal Behavior. Attempts have been made to study the interpersonal behavior of families of schizophrenics *in vivo;* some investigators have gone into the home, others have brought the family into a hospital setting and observed the interaction between family members for an hour or so at a time, others have brought the family into a laboratory setting (National Institute of Mental Health) where the family lives under actual but known conditions for months at a time.

---

[1] The conclusion that the Rorschach protocols of the mothers of schizophrenics were undistinguished from the Rorschach protocols of the mothers of normals (as in the research by Winder & Kantor, 1958, Baxter et al., 1963) is an interpretation of the results made by the present author. In fact, in both of these articles, the authors conclude that there *are* significant differences. However, in the article by Winder and Kantor, the mean rating of the degree of maturity of personality development for the mothers of the schizophrenics was 2.89, for the mothers of the normals, 2.43. In the article by Baxter et al., the means of the ratings of the degree of utilization of psychologically immature defenses on the Rorschach by the parents of poor premorbid schizophrenics, good premorbid schizophrenics, and neurotics are, respectively, 19.43, 19.62, and 19.49. Though in both of these investigations valid statistical significance was demonstrated between the obtained means, the actual means, in both researches, are so similar to each other that the interpretation of *psychologically* significant differences between groups on the basis of the obtained *statistically* significant differences seemed a highly doubtful conclusion.

In the study of the interpersonal relationships in the actual home setting, Behrens and Goldfarb (1958) observed that the personality of the mother seemed to set the tone of the family milieu and that there seemed to be a direct relationship between the degree of pathology that could be seen in the family setting and the degree of psychopathology demonstrated by the child. The homes they observed appeared physically deteriorated and crowded. There was a basic isolation between the mother and father, and the fathers were basically passive. Confusion and disorganization characterized the family atmosphere, with the family demonstrating inadequate mechanisms to handle emotional flareups. The intensive observation of one mother-child interaction (Karon & Rosberg, 1958) yielded the observation that the mother was unempathic. She blocked verbalizations of emotions and tended to live vicariously through the child, but her relationship to the child appeared to involve a basic, though unconscious, hostility and rejection. The mother was an obsessive-compulsive personality, dominated the home, and was unable to accept herself as a woman. The intensive observation of 51 families (Donnelly, 1960) tends to confirm this finding. Observing the mother-child interaction in the home, utilizing the Fels Parent Behavior Scales, Donnelly found that mothers treated a psychotic child differently than their other nonpsychotic children. To the psychotic child, the mother was generally less warm, less accepting, less empathic, more punitive, more controlling, and more overprotective. The father was passive, but more rational than the mother in relation to the child. Psychotic children tended to come from homes characterized as less well adjusted, full of discord, and low in sociability. However, in comparing the family interaction of schizophrenic patients with those of normal controls, both Perr (1958) and Meyers and Goldfarb (1961) found little that could stand as a valid measure of distinction between the two groups of families. Perr found that the parents of schizophrenics tend to show more self-deception and to describe themselves as being more hostile. Meyers and Goldfarb found that the mothers of schizophrenic children appeared less capable of formulating a consistent definition of the world for the child.

A method of directly assessing the interpersonal behavior of husband and wife was introduced by Strodtbeck (1951). He posed questions to each parent individually, then he brought them together and had them discuss those points where their attitudes differed. Farina and his associates (Bell, Garmezy, Farina, & Rodnick, 1960; Farina, 1960; Farina & Dunham, 1963) utilized this method to study the families of schizophrenic patients. The questionnaire they used was the PARI. They found that they could distinguish the interpersonal behavior of schizophrenics otherwise described as having good or poor premorbid adjustment. In these studies, mother dominance was discerned in the families of the poor premorbid group only, with interpersonal conflict greatest in that group. In comparing the family interaction of the schizophrenic patient with those of normal controls, Bell et al. (1960) found that in the family constellation of normals, authority tended to be shared by both parents, and parental conflict was at a minimum, although even here there was a trend towards maternal dominance.

Bishop (1951) reported a method of studying the mother-child interaction under live, yet controlled, conditions. The mother and child were brought into a play room where the interpersonal behavior was observed directly. In 1954, Bowen introduced the principle of this technique to the study of families of schizophrenics. Families were brought into what came to be known as the Family Study Section of NIMH (National Institute of Mental Health), and there they were observed living under actual but known conditions for long periods of time (6 months–2 years). Observations based on families living under these conditions revealed that the mothers of schizophrenics showed extremely domineering, smothering, close relationships with the child (Dworin & Wyant, 1957), with the mothers utilizing threat of deprivation to control the child. Bowen, Dysinger, and Basamanie (1959) observed the presence of marked emotional distance and intense conflict between the parents. The fathers were emotionally immature and unable to define their role in the family and unable to make decisions; the mothers were usually the dominant ones, affecting a close relationship with the child to the exclusion of the father. Brodey (1959) found that the behavior of the families of schizophrenics was characterized by a selective utilization of reality, particularly the use of externalization, and that the interpersonal relationships were highly narcissistic.

Perception of Parental Behavior by Patients. Several studies have indicated that schizophrenics tend to have experienced their mother as having been rejecting (Bolles, Metzger, & Pitts, 1941; Lane & Singer, 1959; Singer, 1954), and dominant, demanding, and overprotective (Garmezy, Clarke, & Stockner, 1961; Heilbrun, 1960; Kohn & Clausen, 1956; McKeown, 1950; Reichard & Tillman, 1950; Schofield & Balian, 1959). However, when one compares the perception of their mothers by normals (Garmezy et al., 1961; Heilbrun, 1960; Lane & Singer, 1959; Singer, 1954) the uniqueness of these attitudes toward the mothers of schizophrenics disappears. Recollections of dominance and overprotectiveness are common for both schizophrenics and normals. Although Heilbrun, Garmezy et al., and Bolles et al. report data which have shown that there is a greater incidence of a feeling of having been rejected on the part of a group of psychiatric patients when compared to medical-surgical controls, the actual incidence of this even in the psychiatric group was only 15% as compared to 1% in the controls. Moreover, Singer and Lane and Singer found that perception of parental relationships during childhood was more a function of the subjects' socioeconomic level than was psychopathology, paralleling a finding by Opler (1957) that familial patterns (parental dominance and attitudes) are a function of cultural factors (Italian versus Irish origin) rather than of psychopathology.

NEUROSIS

As compared to the research in the area of schizophrenia, investigations of the dynamics of the family life of neurotics are few and generally restricted to data gleaned from case histories. From these studies it appears that the neurotic behavior of the child is a direct function of the neurotic behavior of the mother (e.g., Fisher & Mendell, 1956; Ingham, 1949; Sperling, 1949, 1951; Zimmerman, 1930). Neurotic behavior in children has been seen to have been related to maternal overprotection (Holloway, 1931; Jacobsen, 1947; Zimmerman, 1930), maternal domination (Mueller, 1945), maternal rejection (Ingham, 1949; Newell, 1934, 1936; Silberpfennig, 1941), separation from the mother during the first 3 years of life (Bowlby, 1940; Ribble, 1941), and oral deprivation (Childers & Hamil, 1932). Neurotic involvement with the mother, where the mother needs the child for the satisfaction of her own needs and discourages the development of emotional separation between the child and herself, has been associated with the development in the child of psychosomatic disorders (Miller & Baruch, 1950; Sperling, 1949) and school phobia (e.g., Davidson, 1961; Eisenberg, 1958; Estes, Haylett, & Johnson, 1956; Goldberg, 1953; Johnson, Falstein, & Suzurek, 1941; Suttenfield, 1954; Talbot, 1957; van Houten, 1948; Waldfogel, Hahn, & Gardner, 1954; Wilson, 1955).

Neurosis in children has also been associated with such factors in the home as poverty (Brown & Moore, 1944; Holloway, 1931) and broken homes (Ingham, 1949; Madow & Hardy, 1947; Wallace, 1935). Silverman (1935), however, found that 75% of the children from broken homes were essentially "normal"; 16% were described as conduct disorders, and only 9% were classifiable as personality problems.

Of the studies that did not use the case history method of data collection, McKeown (1950) found that neurotic children perceive their mothers as demanding, antagonistic, and setting inordinately high standards for them to meet. Stein (1944) found that neurotics tended to perceive themselves as having been rejected, particularly as compared to the perception of their family life held by normals (Bolles et al., 1941). Although Kundert (1947) found that whether justified by experience or not (e.g., separation due to hospitalization of mother or child), emotionally disturbed children, in general, fear being deserted by their mothers and cling to them compulsively. The Rorschach protocols of mothers of neurotics reveal that they tend to utilize psychological mechanisms which abrogate reality, for example, denial and projection (Baxter et al., 1963).

BEHAVIOR DISORDER

The research on the family background of individuals whose personality problems take the form of antisocial behavior is scanty. Shaw and McKay (1932) found no differences in the incidence of broken homes from cases referred to Cook County Juvenile Court (36%) as compared to a random sample of children in the Chicago public school system (42%). Behavior disorders in children have been seen to have been related to neurotic behavior in their parents (Field, 1940; Huschka, 1941), primarily involving maternal rejection, overt and covert. In line with a social learning hypothesis, another interesting finding is that a correlation has been found between antisocial behavior in children and the children's perception of parents' antisocial behavior (Bender, 1937; K. Friedlander, 1945; Williams, 1932).

DISCUSSION

Let us now summarize what conclusions can be drawn from these data which illuminate the role of the family in the development of psychopathology. As regards the families of schizophrenics, from an overview of the research which has investigated the pattern of parent-child interaction of this pathological group considered without reference to any other pathological or control group, several factors emerge which seem to characterize this group, regardless of the method of data collection, that is, whether by case history, interview, psychological test, or direct observation. Families of schizophrenics seem to be characterized by mothers who are dominant, fathers who are passive, and considerable family disharmony. The mother is overprotective, overpossessive, and overcontrolling, yet basically, albeit unconsciously, rejecting. These mothers frown on sex, are inconsistent in their methods of discipline, and introduce modes of thinking, feeling, and behaving which are not reality oriented. In light of the fact that these patterns emerge as a function of almost all methods of data collection, these results seem very impressive. Had our review of these data stopped here, we would have had apparent verification of the thesis that certain kinds of mother-child relationships and family atmospheres indeed account for the development of schizophrenia in the offspring. However, when each of these parental characteristics is compared with those which emerge from the analysis of the family situation of the normal (apparently nonpsychiatrically-involved) individual, each characteristic that is found to be typical of the families of schizophrenics is found to exist in the families of the controls as well. Furthermore, research which has attempted to make direct comparisons between the families of children in different categories of psychopathology (e.g., Baxter et al., 1963; Fisher et al., 1959; Frazee, 1953; Freeman et al., 1959; D. Friedlander, 1945; Inlow, 1933; McKeown, 1950; Oltman et al., 1952; Pollack & Malzberg, 1940; Pollack, Malzberg, & Fuller, 1936) reveals no significant or consistent differences in the psychological structure of the families.

The results are the same with regard to the families of the neurotics as well. At first glance, it appears that the mother's neurotic involvement with the child is causally associated with the neurotic behavior of the child. However, the essential characteristics of this involvement—maternal overprotectiveness, maternal domination, maternal rejection, deprivation and frustration, and the mothers fostering an almost symbiotic relationship between themselves and their children—are basically the same as those found in the families of schizophrenics and of children with behavior disorders. Moreover, in many respects, it would be hard, on blind analysis, to distinguish the family which produced an emotionally disturbed child from that which produced the so-called normal or well-adjusted child.

It seems apparent that the major conclusion that can be drawn from these data is that there is no such thing as a schizophrenogenic or a neurotogenic mother or family. At least these data do not permit of the description of a particular constellation of psychological events within the home and, in particular, between mother and child that can be isolated as a unique factor in the development of one or the other kind of personality disorder. If one is looking for *the* factor to account for the development of neurosis or schizophrenia, that one factor does not appear to exist as a clear cut finding in the research.

It is incumbent upon us to wonder why the research literature does not permit support of a hypothesis regarding parental influence on the psychological development of children in the manner we hypothesized. One of the major problems with which we must contend is that human behavior is a very complicated event, determined by many factors, and not clearly understood out of the context in which it occurs, and, in this regard, not everyone reacts in a like manner to similar life experiences. For example, strict discipline is reacted to differently when this occurs in a "warm" or "cold" home atmosphere (Sears, 1961); maternal rejection is reacted to differently where the father is accepting and warm (McCord et al., 1961) as well as where the father can be a buffer between the child and the overprotective mother (Witmer, 1933). Emphasizing the multivariate aspect of the determinants of behavior, one notes that Madow and Hardy (1947) reported that of the soldiers who broke down with neurotic reactions there was a high incidence of those coming from broken homes. Amongst those soldiers who did not break down, the incidence of coming from a broken home was 11–15%; the incidence of broken homes in the history of soldiers who did break down was 36%. Statistically, there is a significant difference between these percentages; however, even the 36% datum leaves 64% of the soldiers who broke down *not* coming from a broken home. Huschka (1941) reported that the incidence of neurotic mothers of problem children is high (42%); however, this leaves 58% of the group *not* accounted for by this factor. Brown and Moore (1944) commented that the incidence of excessive poverty, drunkenness, and family conflict in soldiers who broke down was significant, but this accounted for only 20% of the cases. Although between 30% (Canavan & Clark, 1923; Lampron, 1933) and 40% (Huschka, 1941; Lidz & Lidz, 1949) of the children born to mothers who are psychotic become psychotic themselves, these percentages do not account for the majority of children born to these mothers. Indeed, Fanning et al. (1938) found that 43% of the children born to mothers who were psychotic were observed to be making an adequate social and personal adjustment; only 11% of that sample of children was not. It should be noted that only half of the samples of mothers studied by Despert (1938) and Clardy (1951) resembled the traditional pattern of what has come to be known as the "schizophrenogenic mother." Finally, Beres and Obers (1950) observed that there is a wide reaction to an experience of emotional deprivation (in this instance, institutionalization) ranging from the development of a schizoid personality to schizophrenia itself, and including neurotic reactions and character disorders. Indeed, 25% of their sample of children who were brought up in institutions appeared to be making a satisfactory adjustment in spite of this ostensibly devastating experience.

Over and above the complexity of human behavior contributing to the inconclusiveness of the results, one must look at the way in which these data have been collected. It might be that the criterion measure, that is, the diagnosis, did not provide the investigator with meaningful groupings of subjects so that consistent findings *could* be obtained. As regards the method of data collection: Case histories may be inadequate in providing basic data; Information can be gross and/or inaccurate; The informant has to rely on memory, and this memory might be consciously or unconsciously selective, or the informant might not be aware of the import of or feel shame in giving certain data. Yet, despite the many limitations of this mode of data collection, some of the primary research in schizophrenia has utilized this method, and almost all of the data with respect to the family life of neurotics and behavior disorders were gathered in this way. These same limitations apply to data that are gathered when the informant is asked to fill out an attitude questionnaire. Surely the data on parents elicited from the children are susceptible to distortion even when given by normal children, no less those who already tend to consciously or unconsciously confound their perception of reality with fantasy. The psychiatric interview is a much more sensitive procedure than the case history or attitude questionnaire. Either structured or open-ended interviews enable the interviewer to follow up leads and possibly detect where information is being omitted for one reason or another. The problem here, however, is that there is always the possibility that distorted or inaccurate data are gathered by the interviewer, either through the kinds of questions asked, or the perception of the answer or of the individual being interviewed. For example, it is interesting to note that although the majority of psychiatric interviewers experienced the mothers of schizophrenics as matching the model of the schizophrenogenic mother—the dominant, overprotective, but basically rejecting mother who induces inappropriateness of thinking in her children—the psychological test evaluation of mothers of schizophrenics failed to confirm these findings. One explanation for this is that the interviewer, already acquainted with the literature regarding the mother of schizophrenics, anticipating to experience the mothers in terms of the ideas about schizophrenogenic mothers, did, indeed, experience them in that way, whereas a more objective evaluation of the patterns of thinking and feeling of these mothers did not confirm the more subjective impression.

In order to try and avoid the pitfalls inherent in data gleaned through case history or interview, investigators hypothesized that direct observation of the mother-child interaction might yield more valid information. Unfortunately, here, too, limitations inherent in the mode of data collection become apparent. Observations of the mother-child interaction in the home or in an observation room in a hospital or clinic are generally restricted to a limited time segment, for example, 1 hour once a week. This factor, in and of itself, limits the observations to a fairly restricted aspect of the spectrum of the interaction between mother and child. Here, too, the behavior to which the observer is exposed may be influenced by the conscious or unconscious attitudes and motives of the parent being observed. It is not too difficult for the parent to present only that behavior which, for one reason or another, she feels it safe to display and to control the presence of other behaviors. Direct observation of the family for extensive periods of time, that is, months, and under controlled but as natural as possible living conditions (as in the Family Study Section of NIMH) avoids the restrictiveness and overcomes, to one degree or another, the artificiality of the relatively brief observation. However, the family is still aware that they are being observed and may, to one degree or another, be unable to act "natural." Moreover, unless the observations are independently made by several people whose reliability of observation has already been established, they may also be influenced by the *Zeitgeist* and perceive the family as being "schizophrenogenic" whether it is or not, mutually reinforcing each other's expectations. A more pressing consideration in evaluating the validity of these kinds of observations is the fact that the interaction, no matter how natural, takes place after the development of the psychopathology. It is quite possible that the aspects of the interpersonal relationship within the family, or between the mother and child in particular, that eventuated in the development of the patterns of thinking, feeling, and behaving characteristic of the schizophrenic or the neurotic, are no longer present; they may have occurred at a time of the child's life long since past and/or under conditions of intimacy not even accessible to the observer. There is no reason to assume that the etiological factors are still functioning or that they will be available to the trained observer even over the course of 6 months. Of course, it might be that whatever differentiates the psychological existence of the schizophrenic from that of the neurotic or of the normal might be so subtle that it is imperceptible to the participants themselves or even the trained observer and, hence, escape notice. Here, one is reminded of Freud's comment that "the years of childhood of those who are later neurotic need not necessarily differ from those who are later normal except in intensity and distinctness [Freud, 1938 (Orig. publ. 1910), p. 583]."

Theorizing about the etiology of psychopathology has characteristically been of the either/or variety. Nineteenth century scientists sought explanations for neurotic and psychotic disorders in the hereditary background of their patients, working from the assumption that many directly inherited the neurotic or psychotic "illness." On the other hand, the scientist of the twentieth century has sought explanations for psychopathology in the experiential aspect of man's life, in his emotional and interpersonal learning. As with most events in our life, the truth is probably somewhere in between these two positions. Indeed, in spite of the emphasis that is placed on the role of experience in the development of personality in psychoanalysis, Freud did not think, at least as regards the etiology of psychopathology, in such categorically black and white terms. He was able to bridge the gap between the nature-nurture extremes:

> We divide the causes of neurotic disease into those which the individual himself brings with him into life, and those which life bring to him—that is to say, into constitutional and accidental. It is the interaction of these as a rule first gives rise to illness [Freud, 1950b (Orig. publ. 1913), p. 122].

> Let us bear clearly in mind that every human being has acquired, by the combined operation of inherent disposition and of external influences of childhood, a special individuality in the exercise of his capacity to love—that is, in the conditions which he sets up for loving, in the impulses he gratifies by it, and in the aims he sets out to achieve in it. . . . We will here provide against misconceptions and reproaches to the effect that we have denied the importance of the inborn (constitutional) factor because we have emphasized the importance of infantile impressions. Such an accusation arises out of the narrowness with which mankind looks for causes, inasmuch as one single causal factor satisfies him, in spite of the many commonly underlying the face of reality. Psycho-Analysis has said much about the "accidental" component in aetiology and little about the constitutional, but only because it could throw new light upon the former, whereas of the latter it knows no more so far than is already known. We deprecate the assumption of an essential opposition between the two series of aetiological factors; we presume rather a perpetual interchange of both in producing the results observed [Freud, 1950a (Orig. publ. 1912), p. 312].

Other psychoanalysts have followed Freud in the presumption of an inherent, predetermined characteristic functioning of the nervous system of the human organism which determines reactions to stimuli pre- and postnatally (e.g., Greenacre, 1941).

Augmenting the clinical observations of psychoanalysis, one must juxtapose the experimental evidence in psychology which indicates that (*a*) individuals reflect characteristic patterns of autonomic activity which are stable and which are typical of them as individuals (Grossman & Greenberg, 1957; Lacey, 1950; Richmond & Lustman, 1955; Wenger, 1941), (*b*) the characteristic patterns of neural activity are identifiable prenatally and are consistent with the patterns of activity observable postnatally (Richards & Newbery, 1938, (*c*) these characteristic patterns of autonomic activity consistently emerge in a factor of lability and balance in which specific personality factors are consistently highly loaded (Darling, 1940; Eysenck, 1956; Eysenck & Prell, 1951; Theron, 1948; van der Merwe, 1948; van der Merwe & Theron, 1947), (*d*) there is greater similarity of autonomic reactivity between identical twins than fraternal twins or ordinary siblings (Eysenck, 1956; Eysenck & Prell, 1951; Jost & Sontag, 1944), and (*e*) there is a selective influence on personality functioning due to the sex of the individual per se. For example, generally boys outnumber girls 2–1 in being referred for psychological help (Bender, 1937; Wile & Jones, 1937). Sears (1961) found a significant difference in the basic mode of self-reported expression of aggression between boys and girls: Girls appeared higher in socially acceptable forms of aggression and high in anxiety regarding hostility, while boys were significantly higher in aggression that was directed against social control. Sears also found that the more punitive the mother is, the more dependent the son becomes but the less dependent the daughter becomes. Newell (1936) found that maternal rejection affected males more than females: Marked increase in aggressive behavior was noted in the boys who experienced rejection, not so with the females. Baruch & Wilcox (1944) noted that interparental tensions lead to different reactions in boys as compared to girls; in boys, it led to ascendance-submission problems, in girls, to an experience of lack of affection.

We end this survey by concluding that we have not been able to find any unique factors in the family of the schizophrenic which distinguishes it from the family of the neurotic or from the family of controls, who are ostensibly free from evidence of patterns of gross psychopathology. In short, we end by stating that the assumption that the family is *the* factor in the development of personality has not been validated. It is interesting to note that Orlansky (1949), in his review of the literature exploring the relationship between certain childhood experiences, for example, feeding, toilet training, thumbsucking, the degree of tactile stimulation by the mother, etc., upon the development of personality characteristics, was also forced to conclude that the data failed to confirm an invariant relationship between the experience in infancy and the resultant personality. Of course, it might well be that the reality of the family is not the important dimension in determining the child's reactions; rather, it might be the perception of the family members, and this might often have little or no relation to the people as they really are. This would mean, then, that in many instances the important variables in the development of psychopathology might be factors which the child brings to the family, the functioning of the nervous and metabolic systems and the cognitive capacity to integrate stimuli into meaningful perceptual and conceptual schema. Indeed, we are left to wonder, as do the psychoanalysts, whether the proclivity towards fantasy distortion of reality might not be *the* factor in the development of psychopathology, and this proclivity might not be always determined by the child's experiences per se.

Obviously, questions regarding the etiology of patterns of personality behavior which are regarded as pathological, unadaptive, or unadjusted cannot be met with simple answers. Apparently, the factors which play a part in the development of behavior in humans are so complex that it would appear that they almost defy being investigated scientifically and defy one's attempts to draw meaningful generalizations from the exploration which has already been done. It is, of course, conceivable that human behavior is so complex that it cannot be reduced to simple terms or be expected to yield unalterable patterns of occurrences. It might also be that what produces psychopathological reactions in one individual does not in another. All this would be understandable in light of the complexity that is the human being, neurologically as well as socially, but it is unfortunate as regards research endeavors. In 1926, Freud wrote:

> Anxiety is the reaction to danger. One cannot, after all, help suspecting that the reason why the affect of anxiety occupies a unique position in the economy of the mind has something to do with the essential nature of danger. Yet dangers are the common lot of humanity; they are the same for everyone. What we need and cannot lay our fingers on is some factor which will explain why some people are able to subject the affect of anxiety, in spite of its peculiar quality, to the normal workings of the mind, or which decides who is doomed to come to grief over the task [Freud, 1936, p. 64].

We end this review of forty years of research without being able to feel that we are any closer to an answer than was Freud.

REFERENCES

AYER, MARY E., & BERNREUTER, R. G. A study of the relationship between discipline and personality traits in little children. *Journal of Genetic Psychology,* 1937, **50**, 165–170.

BARRY, H. Orphanhood as a factor in psychoses. *Journal of Abnormal and Social Psychology,* 1936, **30**, 431–438.

BARRY, H., & BOUSFIELD, W. A. Incidence of orphanhood among fifteen hundred psychotic patients. *Journal of Genetic Psychology,* 1937, **50**, 198–202.

BARTMEIER, L. H. Deprivations during infancy and their effects upon personality development. *American Journal of Mental Deficiency,* 1952, **56**, 708–711.

BARUCH, DOROTHY W., & WILCOX, J. ANNIE. A study of sex differences in preschool children's adjustment coexistent with inter-parental tensions. *Journal of Genetic Psychology*, 1944, **64**, 281–303.

BAXTER, J. C., BECKER, J., & HOOKS, W. Defensive style in the families of schizophrenics and controls. *Journal of Abnormal and Social Psychology,* 1963, **66,** 512–518.

Behrens, Marjorie L., & Goldfarb, W. A study of patterns of interaction of families of schizophrenic children in residential treatment. *American Journal of Orthopsychiatry,* 1958, **28,** 300–312.

Bell, R. Q., Garmezy, N., Farina, A., & Rodnick, E. H. Direct study of parent-child interaction. *American Journal of Orthopsychiatry,* 1960, **30,** 445–452.

Bender, Lauretta. Reactive psychosis in response to mental disease in the family. *Journal of Nervous and Mental Disease,* 1936, **83,** 143–289.

Bender, Lauretta. Behavior problems in the children of psychotic and criminal parents. *Genetic Psychology Monographs,* 1937, **19,** 229–339.

Beres, D., & Obers, S. J. The effects of extreme deprivation in infancy on psychic structure in adolescence: A study in ego development. *Psychoanalytic Study of the Child,* 1950, **5,** 212–235.

Bishop, Barbara M. Mother-child interaction and the social behavior of children. *Psychological Monographs,* 1951, **65**(11, Whole No. 328).

Bolles, Marjorie M., Metzger, Harriet F., & Pitts, Marjorie W. Early home background and personal adjustment. *American Journal of Orthopsychiatry,* 1941, **11,** 530–534.

Bowen, M., Dysinger, R. H., & Basamanie, Betty. The role of the father in families with a schizophrenic patient. *American Journal of Psychiatry,* 1959, **115,** 1017–1020.

Bowlby, J. The influence of early environment in the development of neurosis and neurotic character. *International Journal of Psychoanalysis,* 1940, **21,** 154–178.

Brodbeck, A. J., & Irwin, O. C. The speech behavior of infants without families. *Child Development,* 1946, **17,** 145–156.

Brodey, W. M. Some family operations in schizophrenia. *Archives of General Psychiatry,* 1959, **1,** 379–402.

Brown, W. T., & Moore, M. Soldiers who break down—family background and past history. *Military Surgeon,* 1944, **94,** 160–163.

Canavan, Myrtelle M., & Clark, Rosamond. The mental health of 463 children from dementia-praecox stock. *Mental Hygiene,* 1923, **7,** 137–148.

Childers, A. T., & Hamil, B. M. Emotional problems in children as related to the duration of breast feeding in infancy. *American Journal of Orthopsychiatry,* 1932, **2,** 134–142.

Clardy, E. R. A study of the development and course of schizophrenic children. *Psychiatric Quarterly,* 1951, **25,** 81–90.

Darling, R. P. Autonomic action in relation to personality traits of children. *Journal of Abnormal and Social Psychology,* 1940, **35,** 246–260.

Davidson, Susannah. School phobia as a manifestation of family disturbance: Its structure and treatment. *Journal of Child Psychology and Psychiatry,* 1961, **1,** 270–287.

Davis, K. Extreme social isolation of a child. *American Journal of Sociology,* 1940, **45,** 554–565.

Despert, Louise J. Schizophrenia in children. *Psychiatric Quarterly,* 1938, **12,** 366–371.

Despert, Louise J. Some considerations relating to the genesis of autistic behavior in children. *American Journal of Orthopsychiatry,* 1951, **21,** 335–350.

Donnelly, Ellen M. The quantitative analysis of parent behavior toward psychotic children and their siblings. *Genetic Psychology Monographs,* 1960, **62,** 331–376.

Dworin, J., & Wyant, O. Authoritarian patterns in mothers of schizophrenics. *Journal of Clinical Psychology,* 1957, **13,** 332–338.

Eisenberg, L. School phobia: A study in the communication of anxiety. *American Journal of Psychiatry,* 1958, **114,** 712–718.

Estes, H. R., Haylett, Clarice H., & Johnson, Adelaide M. Separation anxiety. *American Journal of Psychotherapy,* 1956, **10,** 682–695.

Eysenck, H. J. The inheritance of extraversion-introversion. *Acta Psychologica,* 1956, **12,** 95–110.

Eysenck, H. J., & Prell, D. B. The inheritance of neuroticism: An experimental study. *Journal of Mental Science,* 1951, **97,** 441–465.

Fanning, Aneita, Lehr, Sara, Sherwin, Roberta, & Wilson, Marjorie. The mental health of children of psychotic mothers. *Smith College Studies in Social Work,* 1938, **8,** 291–343.

Farina, A. Patterns of role dominance and conflict in parents of schizophrenic patients. *Journal of Abnormal and Social Psychology,* 1960, **61,** 31–38.

Farina, A., & Dunham, R. M. Measurement of family relationships and their effects. *Archives of General Psychiatry,* 1963, **9,** 64–73.

Field, Minna A. Maternal attitudes found in twenty-five cases of children with primary behavior disorders. *American Journal of Orthopsychiatry,* 1940, **10,** 293–311.

Fisher, S., Boyd, Ina, Walker D., & Sheer, Dianne. Parents of schizophrenics, neurotics, and normals. *Archives of General Psychiatry,* 1959, **1,** 149–166.

Fisher, S., & Mendell, D. The communication of neurotic patterns over two and three generations. *Psychiatry,* 1956, **19,** 41–46.

Fleck, S., Lidz, T., & Cornelison, Alice. Comparison of parent-child relationships of male and female schizophrenic patients. *Archives of General Psychiatry,* 1963, **8,** 1–7.

Frazee, Helen E. Children who later became schizophrenic. *Smith College Studies in Social Work,* 1953, **23,** 125–149.

Freeman, R. V., & Grayson, H. M. Maternal attitudes in schizophrenia. *Journal of Abnormal and Social Psychology,* 1955, **50,** 45–52.

Freeman, H. E., & Simmons, O. G. Mental patients in the community: Family settings and performance levels. *American Sociological Review,* 1958, **23,** 147–154.

Freeman, H. E., Simmons, O. G., & Bergen, B. J. Possessiveness as a characteristic of mothers of schizophrenics. *Journal of Abnormal and Social Psychology,* 1959, **58,** 271–273.

Freud, S. *Inhibitions, symptoms, and anxiety.* London: Hogarth Press, 1936.

Freud, S. Three contributions to the theory of sex. In A. A. Brill (Ed.), *The basic writings of Sigmund Freud.* (Orig. Publ. 1910) New York: Modern Library, 1938. P. 583.

Freud, S. The dynamics of the transference. (Orig. publ. 1912) In, *Collected papers.* Vol. 2. London: Hogarth Press, 1950. Pp. 312–322. (a)

Freud, S. The predisposition to obsessional neurosis. (Orig. publ. 1913) In, *Collected papers.* Vol. 2. London: Hogarth Press, 1950. Pp. 122–132. (b)

Friedlander, D. Personality development of twenty-seven children who later became psychotic. *Journal of Abnormal and Social Psychology,* 1945, **40**, 330–335.

Friedlander, Kate. Formation of the antisocial character. *Psychoanalytic Study of the Child,* 1945, **1**, 189–203.

Garmezy, N., Clarke, A. R., & Stockner, Carol. Child rearing attitudes of mothers and fathers as reported by schizophrenic and normal patients. *Journal of Abnormal and Social Psychology,* 1961, **63**, 176–182.

Gerard, D. L., & Siegal, L. The family background of schizophrenia. *Psychiatric Quarterly,* 1950, **24**, 47–73.

Goldberg, Thelma B. Factors in the development of school phobia. *Smith College Studies in Social Work,* 1953, **23**, 227–248.

Goldfarb, W. The effects of early institutional care on adolescent personality (graphic Rorschach data). *Child Development,* 1943, **14**, 213–223 (a)

Goldfarb, W. Infant rearing and problem behavior. *American Journal of Orthopsychiatry,* 1943, **13**, 249–266. (b)

Goldfarb, W. The effects of early institutional care on adolescent personality. *Journal of Experimental Education,* 1943, **12**, 106–129. (c)

Goldfarb, W. Psychological deprivation in infancy. *American Journal of Psychiatry,* 1945, **102**, 19–33. (a)

Goldfarb, W. Psychological privation in infancy and subsequent adjustment. *American Journal of Orthopsychiatry,* 1945, **15**, 247–255. (b)

Greenacre, Phyllis. The predisposition to anxiety. *Psychoanalytic Quarterly,* 1941, **10**, 66–94.

Grossman, H. J., & Greenberg, N. H. Psychosomatic differentiation in infancy. I. Autonomic activity in the newborn. *Psychosomatic Medicine,* 1957, **19**, 293–306.

Guertin, W. H. Are differences in schizophrenic symptoms related to the mother's avowed attitudes toward child rearing? *Journal of Abnormal and Social Psychology,* 1961, **63**, 440–442.

Hajdu-Grimes, Lilly. Contributions to the etiology of schizophrenia. *Psychoanalytic Review,* 1940, **27**, 421–438.

Hattwick, Berta W. Interrelations between the preschool child's behavior and certain factors in the home. *Child Development,* 1936, **7**, 200–226.

Hattwick, Berta W., & Stowell, Margaret. The relation of parental over-attentiveness to children's work habits and social adjustment in kindergarten and the first six grades of school. *Journal of Educational Research,* 1936, **30**, 169–176.

Heilbrun, A. B. Perception of maternal childbearing attitures in schizophrenics. *Journal of Consulting Psychology,* 1960, **24**, 169–173.

Holloway, Edith. A study of fifty-eight problem children, with emphasis upon the home situation as a causative factor in producing conflict. *Smith College Studies in Social Work,* 1931, **1**, 403.

Huschka, Mabel. Psychopathological disorders in the mother. *Journal of Nervous and Mental Disease,* 1941, **94**, 76–83.

Ingham, H. V. A statistical study of family relationships in psychoneurosis. *American Journal of Psychiatry,* 1949, **106**, 91–98.

Inlow, Ruby S. The home as a factor in the development of the psychosis. *Smith College Studies in Social Work,* 1933, **4**, 153–154.

Jacobsen, Virginia. Influential factors in the outcome of treatment of school phobia. *Smith College Studies in Social Work,* 1947, **19**, 181–202.

Johnson, Adelaide M., Falstein, E. I., Szurek, S. A., & Svendsen, Margaret. School phobia. *American Journal of Orthopsychiatry,* 1941, **11**, 702–711.

Jost, H., & Sontag, L. W. The genetic factor in autonomic nervous system function. *Pyschosomatic Medicine,* 1944, **6**, 308–310.

Karon, B. P., & Rosberg, J. Study of the mother-child relationship in a case of paranoid schizophrenia. *American Journal of Psychotherapy,* 1958, **12**, 522–533.

Kasanin, J., Knight, Elizabeth, & Sage, Priscilla. The parent-child relationship in schizophrenia. *Journal of Nervous and Mental Disease,* 1934, **79**, 249–263.

Kohn, M. L., & Clausen, J. A. Parental authority behavior and schizophrenia. *American Journal of Orthopsychiatry,* 1956, **26**, 297–313.

Kundert, Elizabeth. Fear of desertion by mother. *American Journal of Orthopsychiatry,* 1947, **17**, 326–336.

Lacey, J. I. Individual differences in somatic response patterns. *Journal of Comparative and Physiological Psychology,* 1950, **43**, 338–350.

Lampron, Edna M. Children of schizophrenic parents. *Mental Hygiene,* 1933, **17**, 82–91.

Lane, R. C., & Singer, J. L. Familial attitudes in paranoid schizophrenics and normals from two socioeconomic classes. *Journal of Abnormal and Social Psychology,* 1959, **59**, 328–339.

Levy, D. M. Maternal overprotection and rejection. *Archives of Neurology and Psychiatry,* 1931, **25**, 886–889.

Levy, D. M. On the problem of delinquency. *American Journal of Orthopsychiatry,* 1932, **2**, 197–211.

Levy, D. M. Primary affect hunger. *American Journal of Psychiatry,* 1937, **94**, 643–652.

Levy, D. M. *Maternal overprotection.* New York: Columbia Univer. Press, 1943.

Lidz, T., Cornelison, Alice R., Fleck, S., & Terry, Dorothy. The intrafamilial environment of the schizophrenic patient: I. The father. *Psychiatry,* 1957, **20,** 329–342. (a)

Lidz, T., Cornelison, Alice R., Fleck, S., & Terry, Dorothy. The intrafamilial environment of schizophrenic patients: II. Marital schism and marital skew. *American Journal of Psychiatry,* 1957, **114,** 241–248. (b)

Lidz, T., Cornelison, Alice R., Fleck, S., & Terry, Dorothy. The intrafamilial environment of the schizophrenic patient. *Psychiatry,* 1957, **20,** 329–342. (c)

Lidz, T., Cornelison, Alice, Terry, Dorothy, & Fleck, S. Intrafamilial environment of the schizophrenic patient: VI. The transmisison of irrationality. *Archives of Neurology and Psychiatry,* 1958, **79,** 305–316.

Lidz, Ruth W., & Lidz, T. The family environment of schizophrenic patients. *American Journal of Psychiatry,* 1949, **106,** 332–345.

Lidz, T., Parker, Neulah, & Cornelison, Alice. The role of the father in the family environment of the schizophrenic patient. *American Journal of Psychiatry,* 1956, **113,** 126–137.

Lowrey, L. G. Personality distortion and early institutional care. *American Journal of Orthopsychiatry,* 1940, **10,** 576–585.

Madow, L., & Hardy, S. E. Incidence and analysis of the broken family in the background of neurosis. *American Journal of Orthopsychiatry,* 1947, **17,** 521–528.

Mark, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *Journal of Abnormal and Social Psychology,* 1953, **48,** 185–189.

McCord, W., McCord, Joan, & Howard, A. Familial correlates of aggression in nondelinquent male children. *Journal of Abnormal and Social Psychology,* 1961, **62,** 79–93.

McKeown, J. E. The behavior of parents of schizophrenic, neurotic, and normal children. *American Journal of Sociology,* 1950, **56,** 175–179.

Meyers, D. I., & Goldfarb, W. Studies of perplexity in mothers of schizophrenic children. *American Journal of Orthopsychiatry,* 1961, **31,** 551–564.

Meyers, D., & Goldfarb, W. Psychiatric appraisals of parents and siblings of schizophrenic children. *American Journal of Psychiatry,* 1962, **118,** 902–908.

Miller, H., & Baruch, D. A study of hostility in allergic children. *American Journal of Orthopsychiatry,* 1950, **20,** 506–519.

Mueller, Dorothy D. Paternal domination: Its influence on child guidance results. *Smith College Studies in Social Work,* 1945, **15,** 184–215.

Newell, H. W. The psycho-dynamics of maternal rejection. *American Journal of Orthopsychiatry,* 1934, **4,** 387–401.

Newell, H. W. A further study of maternal rejection. *American Journal of Orthopsychiatry,* 1936, **6,** 576–589.

Oltman, Jane E., McGarry, J. J., & Friedman, S. Parental deprivation and the "broken home" in dementia praecox and other mental disorders. *American Journal of Psychiatry,* 1952, **108,** 685–694.

Opler, M. K. Schizophrenia and culture. *Scientific American,* 1957, **197,** 103–110.

Orlansky, H. Infant care and personality. *Psychological Bulletin,* 1949, **46,** 1–48.

Pearson, G. H. Some early factors in the formation of personality. *American Journal of Orthopsychiatry,* 1931, **1,** 284–291.

Perr, H. M. Criteria distinguishing parents of schizophrenic and normal children. *Archives of Neurology and Psychiatry,* 1958, **79,** 217–224.

Plank, R. The family constellation of a group of schizophrenic patients. *American Journal of Orthopsychiatry,* 1953, **23,** 817–825.

Pollock, H. M., & Malzberg, B. Hereditary and environmental factors in the causation of manic-depressive psychoses and dementia praecox. *American Journal of Psychiatry,* 1940, **96,** 1227–1244.

Pollock, H. M., Malzberg, B., & Fuller, R. G. Hereditary and environmental factors in the causation of dementia praecox and manic-depressive psychoses. *Psychiatric Quarterly,* 1936, **10,** 495–509.

Preston, G. H., & Antin, Rosemary. A study of children of psychotic parents. *American Journal of Orthopsychiatry,* 1932, **2,** 231–241.

Prout, C. T., & White, Mary A. A controlled study of personality relationships in mothers of schizophrenic male patients. *American Journal of Psychiatry,* 1950, **107,** 251–256.

Reichard, Suzanne, & Tillman, C. Patterns of parent-child relationships in schizophrenia. *Psychiatry,* 1950, **13,** 247–257.

Ribble, Margarethe A. Disorganizing factors of infant personality. *American Journal of Psychiatry,* 1941, **98,** 459–463.

Richards, T. W., & Newbery, Helen. Studies in fetal behavior: III. Can performance on test items at six months postnatally be predicted on the basis of fetal activity? *Child Development,* 1938, **9,** 79–86.

Richmond, J. B., & Lustman, S. L. Autonomic function in the neonate: I. Implications for psychosomatic theory. *Psychosomatic Medicine,* 1955, **17,** 269–275.

Schofield, W., & Balian, L. A comparative study of the personal histories of schizophrenic and nonpsychiatric patients. *Journal of Abnormal and Social Psychology,* 1959, **59,** 216–225.

Sears, R. R. Relation of early socialization experiences to aggression in middle childhood. *Journal of Abnormal and Social Psychology,* 1961, **63,** 466–492.

Sears, R. R., Whiting, J. W. M., Nowlis, V., & Sears, Pauline S. Some child-rearing antecedents of aggression and dependency in young children. *Genetic Psychology Monographs,* 1953, **47,** 133–234.

Shaw, C. R., & McKay, H. D. Are broken homes a causative factor in juvenile delinquency? *Social Forces,* 1932, **10**, 514–524.

Shoben, E. J. The assessment of parental attitudes in relation to child adjustment. *Genetic Psychology Monographs,* 1949, **39**, 101–148.

Silberpfennig, Judith. Mother types encountered in child guidance clinics. *American Journal of Orthopsychiatry,* 1941, **11**, 475–484.

Silverman, B. The behavior of children from broken homes. *American Journal of Orthopsychiatry,* 1935, **5**, 11–18.

Singer, J. L. Projected familial attitudes as a function of socioeconomic status and psychopathology. *Journal of Consulting Psychology,* 1954, **18**, 99–104.

Sperling, Melitta. The role of the mother in psychosomatic disorders in children. *Psychosomatic Medicine,* 1949, **11**, 377–385.

Sperling, Melitta. The neurotic child and his mother: A psychoanalytic study. *American Journal of Orthopsychiatry,* 1951, **21**, 351–362.

Spitz, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanalytic Study of the Child,* 1945, **1**, 53–74.

Stein, Lucille H. A study of over-inhibited and unsocialized-aggressive children. *Smith College Studies in Social Work,* 1944, **15**, 124–125.

Stringer, Joyce R. Case studies of the families of schizophrenics. *Smith College Studies in Social Work,* 1962, **32**, 118–148.

Strodtbeck, F. L. Husband-wife interaction over revealed differences. *American Sociological Review,* 1951, **16**, 468–473.

Suttenfield, Virginia. School phobia: A study of five cases. *American Journal of Orthopsychiatry,* 1954, **24**, 368–380.

Talbot, Mira. School Phobia: A workshop: I. Panic in school phobia. *American Journal of Orthopsychiatry,* 1957, **27**, 286–295.

Theron, P. A. Peripheral vasomotor reaction as indices of basic emotional tension and lability. *Psychosomatic Medicine,* 1948, **10**, 335–346.

Tietze, Trude. A study of mothers of schizophrenic patients. *Psychiatry,* 1949, **12**, 55–65.

van der Merwe, A. B. The diagnostic value of peripheral vasomotor reactions in the psychoneuroses. *Psychosomatic Medicine,* 1948, **10**, 347–354.

van der Merwe, A. B., & Theron, P. A. A new method of measuring emotional stability. *Journal of General Psychology,* 1947, **37**, 109–124.

van Houten, Janny. Mother-child relationships in twelve cases of school phobia. *Smith College Studies in Social Work,* 1948, **18**, 161–180.

Wahl, C. W. Some antecedent factors in the family histories of 392 schizophrenics. *American Journal of Psychiatry,* 1954, **110**, 668–676.

Wahl, C. W. Some antecedent factors in the family histories of 568 male schizophrenics of the United States Navy. *American Journal of Psychiatry,* 1956, **113**, 201–210.

Waldfogel, S., Hahn, Pauline B., & Gardner, G. E. A study of school phobia in children. *Journal of Nervous and Mental Disease,* 1954, **120**, 399.

Wallace, Ramona. A study of the relationship between emotional tone of the home and adjustment status in cases referred to a travelling child guidance clinic. *Journal of Juvenile Research,* 1935, **19**, 205–220.

Walters, Jean H. A study of the family relationships of schizophrenic patients. *Smith College Studies in Social Work,* 1939, **9**, 189–191.

Watson, G. A. A comparison of the effects of lax versus strict home training. *Journal of Social Psychology,* 1934, **5**, 102–105.

Wenger, M. A. The measurement of individual differences in autonomic balance. *Psychosomatic Medicine,* 1941, **3**, 427–434.

Wile, I. S., & Jones, Ann B. Ordinal position and the behavior disorders of young children. *Journal of Genetic Psychology,* 1937, **51**, 61–93.

Williams, H. D. Causes of social maladjustment in children. *Psychological Monographs,* 1932, **43**(**1**, Whole No. 194)

Wilson, Margaret J. Grandmother, mother, and daughter in cases of school phobia. *Smith College Studies in Social Work,* 1955, **25**, 56–57.

Winder, C. L., & Kantor, R. E. Rorschach maturity scores of the mothers of schizophrenics. *Journal of Consulting Psychology,* 1958, **22**, 438–440.

Witmer, Helen L. Parental behavior as an index to the probable outcome of treatment in a child guidance clinic. *American Journal of Orthopsychiatry,* 1933, **3**, 431–444.

Zimmerman, Anna C. Parental adjustments and attitudes in relation to the problems of five- and six-year-old children. *Smith College Studies in Social Work,* 1930, **1**, 406–407.

Zuckerman, M., Oltean, Mary, & Monashkin, I. The parental attitudes of mothers of schizophrenics. *Journal of Consulting Psychology,* 1958, **22**, 307–310.

(Received February 24, 1965)

Psychological Bulletin
1965, Vol. 64, No. 3, 206–216

FUNCTIONAL AND ADAPTATIVE ASPECTS OF CELLULAR REGULATORY MECHANISMS

W. W. MEISSNER
Harvard Medical School

Recent developments in biochemical genetics, namely, the understanding of the mechanisms of regulation of cellular processes through the operation of regulator genes, brings the genome into intimate involvement in the adaptative processes of the organism. The dichotomy of structure and function is beginning to dissolve in a deeper understanding of the function of the organism. Available evidence indicates that similar regulatory mechanisms are operative in man which control the integration of hormonal activity and serve as the organic basis for the correlation of structure and function, even in higher cerebral functions related to emotional and learning behavior.

The deepening scientific understanding of basic cellular processes has reached a point in recent years where an understanding of such subcellular mechanisms can begin to contribute to our deeper understanding of higher functions of the organism. The currently developing understanding of basic processes has opened avenues of thinking and research which strike much more closely at the heart of the more strictly functional aspects of the behaving organism. In some radical sense, the lines of division between structure and function are beginning to fade. The possibility presents itself on many fronts of linking various aspects of function with basic cellular processes. It may be possible to reach some understanding of the effect of functional adaptations of the organism, like learning, or functional maladaptations, like schizophrenia, in relation to basic processes.

One of the most significant areas in which our understanding has approached this level is in the exploration of cellular regulatory mechanisms. A new understanding of the relation between cellular functioning and structural change has been developed in the last few years, based on an appreciation of the role of genetic mechanisms in controlling cellular responses. Much of the basic understanding of these mechanisms was achieved in simple cellular systems like the bacteria, most notably *Escherichia coli.* A review of these cellular mechanisms, therefore, will help to provide a clearer understanding of their relevance and significance for more strictly functional and specifically psychological processes.

Over the course of the many years in which students of human behavior have been investigating genetic aspects of behavior, their general orientation toward their subject matter was principally dictated by the lines of thought laid down by the science of genetics itself. Under the impact of Mendelian genetics, psychologists and psychiatrists concentrated their efforts on the determination of the genotypes corresponding to the psychological phenotypes. The overriding concern was to identify the genetic factors to which the observed variance in psychological abilities and performance could be ascribed. The formulation of the notion of inborn errors of metabolism (Garrod, 1923) and the establishment of a link between genic determination and enzyme synthesis gave the stimulus to the tremendous effort to specify genetic factors in the etiology of mental illness.

These are the interests which dominate contemporary thinking about behavioral genetics. A common factor in this general approach has been the operative conception of the genetic material as a hereditary given which determines the basic structure of the organism. The genetically determined structure is then regarded as capable of being modified to one degree or other, or in one fashion or other, by a variety of environmental influences. This basic conception of the gene as somehow related to structure has underlain the whole course of our thinking about the relationship between heredity and environment. The hereditary (genetic) factor was regarded as responsible for the structure of the organism, whether macroscopically in terms of gross architecture of the body, or microscopically in terms of cell structure or even the molecular structure of enzyme systems. The given hereditary structure was then modified developmentally and functionally by environmental influences. Consequently, it was possible to approach the study of the interaction between heredity and environment in terms of the amount of variance which could be attributed to each.

The identification of deoxyribose nucleic acid (DNA) as the genetic material (Avery, MacLeod, & McCarthy, 1944) gave rise to a new era of biochemical genetics. With the elucidation of the structure of DNA by Watson and Crick (1953), a tremendous impetus was given to the study of the biochemical bases for the transfer of genetic information. Little by little the picture of the mechanisms by which the genetic information was transferred from DNA in the nuclear chromatin to specific proteins in the cellular cytoplasm was pieced together. The role of transfer ribonucleic acid (t-RNA) in protein synthesis (Hoagland, Stephenson, Scott, Hecht, & Zamecnik, 1958), the clarification of the role of messenger ribonucleic acid (m-RNA) in protein synthesis (Brenner, Jacob, & Meselson, 1961; Nirenberg, & Mathaei, 1961), and the slow progress in breaking the genetic code (Crick, Barnett, Brenner, & Watts-Tobin, 1961; Jones & Nirenberg, 1962) has given us a consistent and increasingly more accurate idea of how the cell produces proteins out of the supply of amino-acids available to it. A clear and concise recounting of this whole development has recently been provided by Beadle (1964). The whole development was directed toward the elucidation, on the molecular and biochemical level, of the mechanisms inherent in the dominant structural orientation of scientific genetics. The DNA template was still more or less regarded as a structural given which served, by the intermediates of m-RNA in conjunction with the RNA in the ribosomal fraction, to give rise to the structural units out of which the organism was constructed, the proteins. To this extent, then, recent developments in biochemical genetics have not altered the basic structural-functional dichotomy which has governed our thinking about behavioral genetics.

While rapid strides were being made in investigating the function of DNA and RNA, other developments were being made which seem quite likely to revolutionize our thinking about the operation of genetic factors in the adaptation of the organism. As soon as it had become clear that DNA was directly involved in protein synthesis, attention was inevitably directed to the question of the regulation of protein synthesis. For it was quite clear that if all of the genes in a given cell were to produce protein at about the same rate, an extremely unwieldy situation would soon develop. The cell in fact varies the rate of protein production under differing conditions. The continued production of unneeded proteins would prove to be very uneconomical (Zalokar, 1961). Furthermore, when the necessity arises to produce a given protein in greater amounts than usual, it is unreasonable to accelerate the total production of RNA in order to produce a sufficient amount of the RNA needed for the production of the one protein. Moreover, the organism develops from a single set of diploid chromosomes in the zygote. Yet in the adult animal tremendous ranges of differentiation are manifest, both in structure and function of the respective cells. Clearly, some mechanism must be provided whose function is to regulate the sequential activation of the gene potential during the different phases of the organism's life.

Once the problem of cellular regulation had been brought into focus, it became abundantly clear that the cell does in fact alter the rates at which various protein products are produced. Most of the studies of cellular control mechanisms have been conducted with bacteria, primarily the ubiquitous *Escherichia coli*. It was found, for example, that aspartate transcarbamylase of *E. coli* was strongly inhibited by cytidine triphosphate, its biosynthetic product. The reactions can be represented in schematic form:

$$\begin{tikzcd} \text{Carbamyl-phosphate} + \text{aspartate} \arrow[r] & \text{carbamyl-aspartate} \arrow[r] & {} \arrow[r] & {} \arrow[r] & \text{Cytidine triphosphate} \arrow[llll, dashed, bend left] \end{tikzcd}$$

Maximal inhibition was achieved with cytidine triphosphate rather than cytidine diphosphate or any other closely related compound, indicating that the inhibition was quite specific. The inhibition was reversed by increasing the concentration of aspartate (Gerhart & Pardee, 1962). A similar reaction pattern was discovered for inhibition of L-threonine deaminase by L-threonine (Changeaux, 1961). Both series of reactions had certain common characteristics: (*a*) the inhibited enzyme was the first in the series which led specifically to the production of the endproduct; (*b*) inhibition was apparently achieved by the endproduct itself, rather than by some derivative; (*c*) when the level of the endproduct decreased, the enzyme proportionally increased the rate of production (reversible regulation). Umbarger (1961) has proposed two models for this sort of feedback inhibition: competitive inhibition between substrate and endproduct with catalytic and inhibitory sites overlapping; or combination of the endproduct with the enzyme at a site which does not interfere with the binding between enzyme and substrate. In the latter case, inhibition of catalysis would be effected by some sort of allosteric effect of the binding of the endproduct to the enzyme. A considerable number of such feedback inhibitions of enzyme systems have been demonstrated (Umbarger, 1961).

More significantly, however, other systems have been discovered which involve the selective activation of enzyme systems and the production of proteins. The shift from the utilization of glucose to the utilization of galactose in the nutrient medium by *E. coli* was found to involve the production of a new protein, β-galactosidase (Pardee, Jacob, & Monod, 1958, 1959). The increase in activity in the aspartate transcarbamylase system was also found to be accompanied by selective synthesis of a new protein having the same electrophoretic mobility as the transcarbamylase (Shepherdson & Pardee, 1960a, 1960b; Yates & Pardee, 1957). Similar effects were also demonstrated for the ornithine transcarbamylase system (Rogers & Novelli, 1959) and for alkaline phosphatase (Torriani, 1960). In each such case of selective activation of the enzyme system, it was shown that the production of new enzyme was in question. But how was the production of a specific enzyme brought about?

There is good evidence to suggest that adaptational effects within bacterial cells affect specific proteins selectively, and that the enzymatic adaptation involves the production of protein from available aminoacids (Hogness, Cohen, & Monod, 1955; Rotman & Spiegelman, 1954). The regulation of enzyme production must in some sense be responsive to the functional state of the cell. Consequently, the activation signal must derive from the cytoplasm or its environment. Thus, the presence of galactose in the bacterial cell acts as a stimulus to the activation of the production of β-galactosidase, and the presence of glutamate in the cell acts as the stimulus for the production of ornithine transcarbamylase. Such substances have come to be known as "inducers" since their action is the induction of enzyme synthesis.

In the DNA-RNA schema for protein production, the site of operation of such inducers can be either the secondary template of the polyribosomes or the primary template of DNA itself. It is quite possible that in different organisms and different systems, one or the other site is affected. But there is good evidence that inducers (and their opposites, the repressors) are influencing the production of protein at the level of the DNA template. Such evidence is based on results of continuous cultivation experiments under repressive conditions (Cocito & Vogel, 1959), on studies of coordinate repression (Ames & Garry, 1959), on studies of the immunity of lysogenic cells (Jacob, 1960), and on investigations into the function of operator genes (Jacob, Perrin, Sanchez, & Monod, 1960). Some convincing evidence has been put forth to suggest that in at least some systems the significant exchange which reflects the effects of induction takes place between the DNA-template and the messenger-RNA. If induction takes place at the RNA level, then RNA must be continuously present in the cell, and induction could be carried out without any RNA synthesis. Kramer and Straub (1956) found that the destruction of existing RNA by ribonuclease did not affect inducibility. Application of ribonuclease after the inducer, however, did significantly impair the production of new enzyme. It has also been found that prevention of induction was possible by the use of analogs to inhibit RNA synthesis (Spiegelman, Halvoørsen, & Ben-Ishai, 1955). Furthermore, if RNA synthesis is the primary effect of induction, the action of DNA is necessarily involved. Destruction of bacterial DNA after the incorporation of P^{32} destroys the cell's ability to produce the induced enzyme (McFall, Pardee, & Stent, 1958). Also Gale and Folkes (1955) were able to induce β-galactosidase formation in a homogenized preparation of *E. coli*; they were able to suppress synthesis of the enzyme with deoxyribonuclease, and subsequently reactivate the process by the addition of bacterial DNA. Synthesis of RNA appears to be a central aspect of the inductive activation of enzyme systems. It has been argued, moreover, that the incorporation of precursors into RNA is a process that continues throughout the growth cycle of the cell (Lark & Lark, 1960) and that the rate of synthesis of ribosomal RNA is more highly responsive to the chemical environment than the synthesis of either proteins or DNA (Neidhardt & Fraenkel, 1961). It provides a very sensitive mechanism of response to slight changes in the environment of the cell (Schaechter, 1961).

The available evidence, of which there is considerably more than we have undertaken to present here, has been formulated into a theory of cellular regulation by Jacob and Monod (1961a, 1961b) and by others. We can gain an insight into the nature of regulatory processes by a brief review of the major points in the theory.

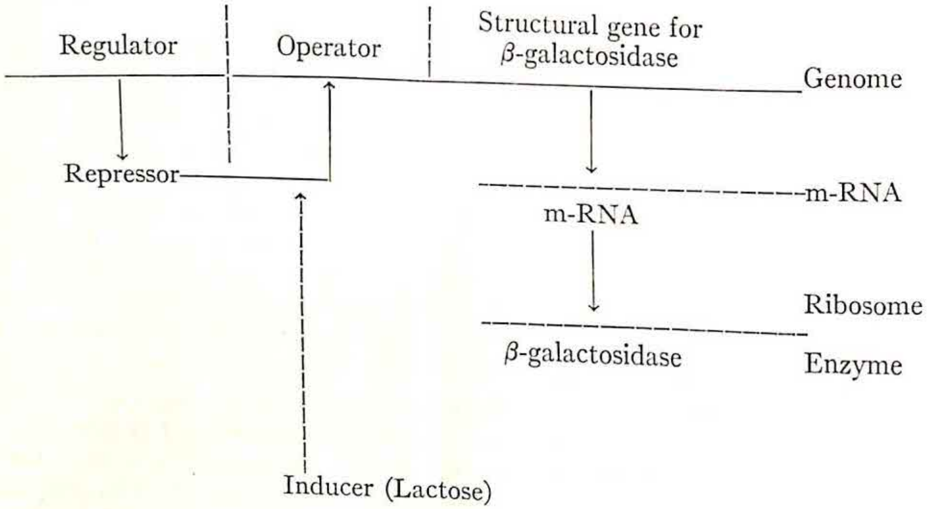
According to this model of cellular regulation, the synthesis of enzymes is subject to a double genetic control. The structural genes, which have long been the objects of genetic investigation, determine the molecular organization of protein products by specifying the sequence of aminoacid residues composing the protein structure. Other functionally specialized genetic determinants are necessary, however, to determine the mode of operation of the genetic mechanism. These specialized genetic components are the regulator and operator genes, and their function is to control the rate of protein synthesis through the intermediate action of cytoplasmic components. Umbarger (1961) has offered the following observation regarding these genetic components:

> A feature of the chain of events between gene and enzyme that has received emphasis in this symposium is the concept that the presence of a particular region of DNA containing the information for the amino acid sequence of a particular protein is not by itself sufficient condition that a particular protein will be formed. The role of the regulator and operator as genetically controlled mechanisms enabling the structural gene to respond to environmental changes has been sufficiently well documented that any attempt to explain the events of gene expression must incorporate these control features into the proposed scheme [p. 301].

Regulator and operator genes exercise their effects through the action of cytoplasmic repressors. The repressor can either be inactivated by the process of induction, or it can be activated by the opposed process of repression; in each instance, the induction or repression is carried out by means of special metabolites.

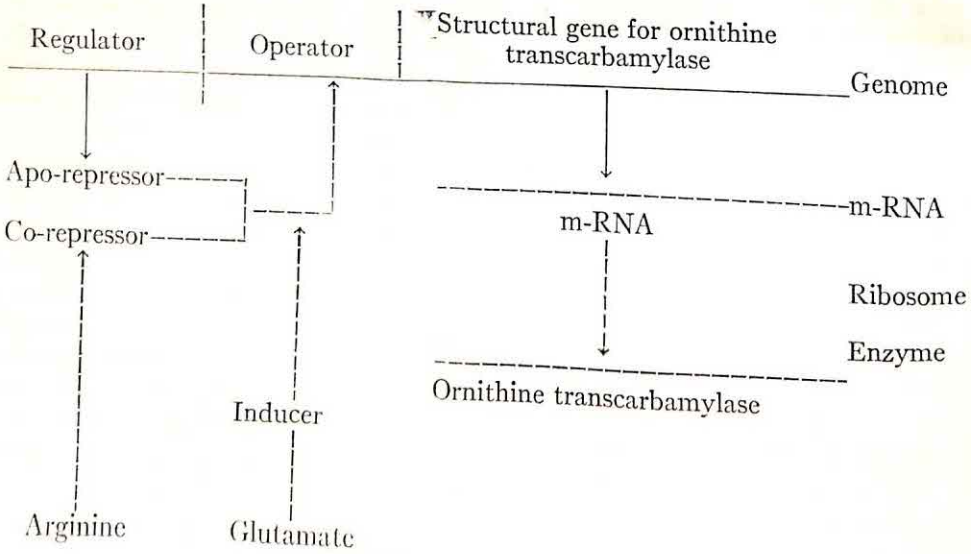
The system operates directly at the level of the synthesis of m-RNA from the DNA gene-template. The messenger is produced by a replicative sequential process from the structural gene. Thus the structural gene transfers the encoded genetic information to the m-RNA, which then transfers the information to the cellular ribosomes determining them to produce the specific protein required by the genetic code. The transcription of the encoding sequence of the structural gene is dependent on an initiating point, the operator gene. Each messenger may code for one or more polypeptides, and a series of structural genes involved in the encoding for a series of interrelated proteins may be under the coordinate control of a single operator. Such a unit of transcription is called an operon. The operator combines specifically and reversibly with the RNA fraction or protein which possesses the proper complementary sequence of amino acids. The availability of the operator locus is essential to the formation of m-RNA, so that the combination of a repressor substance with the operator locus can block the cytoplasmic transcription of m-RNA, and thereby diminish the rate of protein synthesis, since the rate of m-RNA synthesis is the critical factor in determining the rate of protein production. The function of the regulator gene in this schema is to synthesize the repressor which combines with and represses the operator locus.

As an example of the operation of this schema we can consider the induction of β-galactosidase by lactose in the cell medium (see above.) The induction process is visualizable in diagram form.



In this case, the repressor substance, produced by the regulator gene, inhibits the action of the operator gene and thus prevents the transcription of m-RNA for the production of β-galactosidase. Absorption of the substrate lactose brings the inducer, lactose, into contact with the repressor. The inducer inactivates the repressor and permits the operator to initiate the transcription of m-RNA. The cell thus responds to the presence of lactose in the cell medium by the initiation of β-galactosidase formation which enables the cell to utilize lactose and galactose as a nutrient.

A similar example can be found in the case of the production of ornithine transcarbamylase.



In this instance, arginine is required as a co-repressor in order that the apo-repressor produced by the regulator gene be effective. Consequently, when the level of arginine in the cell is low, enzyme synthesis is still possible. Synthesis of the enzyme, however, increases the action of the enzyme system on the substrate glutamate, thus increasing the concentration of the product arginine. This is an elaboration of the sort of endproduct inhibition we have already discussed above. Glutamate, in addition, acts as an inducer, stimulating the production of enzyme by inhibiting the combined action of apo-repressor and co-repressor on the operator.

These systems are only two of a number that have been defined with some security. Consequently, they are representative only of a family of such intracellular control mechanisms, the full scope of which is more than likely quite broad and varied. Moreover, as Jacob and Monod (1961a, 1961b) have pointed out, not all the points of these models can be asserted with the same confidence. The most firmly established factor is the existence of regulator genes. This conclusion is supported by mutations of structural genes which affect the alteration of protein structure yet leave the regulatory mechanism unaltered, conversely, by mutations which alter the regulatory mechanism but do not alter the protein structure. It also seems clear that regulator genes can by means of specific cytoplasmic substances inhibit the expression of structural genes. This can be established by differential properties exhibited by genetically identical zygotes which depend on the origin of the cytoplasm, and by the fact that the absence of the regulator gene or its product results in uncontrolled synthesis of the protein at maximal rates. The existence of the operator as the site of action of the repressor is deduced from the specificity of repressor action, and the identification of the operator with a genetic segment is suggested by the observation that a single operator gene may control the expression of several adjacent structural genes. The assumptions concerning the interaction of the repressors with inducers or with co-repressors are the weakest part of the model (Jacob & Monod, 1961a, 1961b).

With regard to this latter point, these same workers have suggested that the available evidence points to the conclusion that the active product of the regulator gene may in fact be a protein substance, rather than the RNA product originally proposed. Several lines of evidence support this conclusion: (*a*) The repressor action has a high degree of stereospecificity; (*b*) the interaction with the operator is virtually immediate and reversible; (*c*) genetic evidence shows that a single gene, and therefore, presumably a single specific macromolecular constituent (the repressor) is responsible for the specificity of the interaction; and (*d*) single mutations of this gene abolish the repressor-inducer interaction while preserving the repressor-operator affinity (Jacob & Monod, 1963). Consequently, Jacob and Monod feel that all properties are accounted for if the repressor is an allosteric protein which possesses two sites, one of which interacts with and binds the operator, the other of which interacts with the inducer. Consequently, the specific effects of small molecule metabolites in activating or inhibiting the transcription of m-RNA at the gentic level are mediated by an allosteric transition of the repressor (Monod, Changeaux, & Jacob, 1963).

The picture which emerges from these fragments of fact and theory gives us the basis for an entirely different understanding of the activity of the genetic material. Through the mechanisms mediated by the regulator and operator genes, the genetic apparatus is brought into continual interaction with the functioning organism. Rather than considering the genes in a strictly structural orientation, we must begin to think of them as continually and dynamically operating to regulate the adaptation of the organism to its environment.

Considerable care is called for in trying to extrapolate present evidence about the genetic regulation of cell processes to the functioning of the more highly differentiated and more highly organized animal organisms. Cellular regulatory mechanisms defined in microorganisms cannot be simply transferred to more complex organisms. Higher organisms must be expected to employ a considerably wider range of regulatory mechanisms than are available to the microorganism. Yet it is also altogether likely that the basic mechanisms defined in the study of *E. coli* can also be found operating at the cellular level of adjustment in more differentiated organisms.

Some evidence has already been accumulated which suggests that similar mechanisms are at work in higher organisms. In the endocrine system, it is well established that endocrine glands can be activated to increase the rate of production of their hormonal products. The increase of the rate of production of some hormones may be an effect of genetic control through the activated production of m-RNA (Karlson, 1962). Production of protein hormones like ACTH might be direct, and production of nonprotein horomones like the steroids might be mediated by the increased production of enzyme systems. The thyroid gland, for example, produces thyroxin and thyroglobulin. The rate of production of these hormones is elevated by the production of the thyroid stimulating hormone (TSH) from the adenohypophysis. Under the stimulus of TSH, the thyroid is activated to increase its production. The epithelial cells of the follicles increase in size, the cell nuclei and nucleoli enlarge and become more prominent, and the endoplasmic reticulum shows a definite increase. Moreover, active RNA production can be demonstrated by direct measurement (Matovinovic & Vickery, 1959). This increased thyroid activity increases the level of thyroxin in the blood which has a feedback effect on the pituitary, suppressing the production of TSH. TSH is produced by certain basophilic cells in the pituitary which are rich in RNA. After suppression by thyroxin, the percentage of basophilic cells decreases. When the level of thyroxin in the blood drops and TSH production is stimulated, the percentage of such basophilic cells markedly increases (Griesbach & Purves, 1943). The basophils of the stimulated hypophysis increase in size and the nuclei become enlarged, showing several prominent nucleoli (Stevens, D'Angelo, Paschkis, Canterow, & Sunderman, 1955). This would suggest that TSH acts in the manner of an inducer on the hormone producing mechanisms of the thyroid gland, increasing the production of thyroid hormones. Similarly, thyroxin has a feedback inhibition effect on the basophilic cells of the pituitary. This latter effect may be mediated by cytoplasmic factors involving inhibition of the enzyme system. It is not clear whether the site of inhibitory action is directly on the basophil cells themselves, or on neurosecretory cells in the hypothalamus. But even here, morphological indications at least suggest that new protein synthesis is going on, and it is quite possible that genetic regulatory mechanisms are involved. Similarly, in the thyroid cells it is quite likely that TSH is acting through genetically regulated synthetic processes.

A strikingly similar case can be made for many of the hormonal interactions which regulated the body economy. Pituitary ACTH stimulates the production of adrenocortical steroids, and the cortisol thus produced has an inhibiting effect on the hypothalamo-pituitary sites of ACTH production. A similar interaction takes place between the gonadotropic hormones and the sex hormones, between luteinizing hormone and the products of the developing corpus luteum, etc. There is good reason, therefore, to think that hormonal action is in many instances directed to the genetic sites of regulation of hormone synthesis, or at least of the regulation of the synthesis of enzymes involved in hormone production. The hormones thus produced from the regulated glands feed back to repress genetic sites responsible for the synthesis of tropic hormones. The frequently observed specificity of target glands might be understood in terms of the specificity of the active sites involved (Zalokar, 1961).

The understanding of the mechanisms of action of hormonal control mechanisms is enhanced by the work of Knox and others (Knox, & Auerbach, 1955; Knox, Auerbach, & Lin, 1956). They have shown that the injection of the substrate tryptophan and of cortisone produced a similar increase in the activity of tryptophan pyrrolase in rats. A similar effect was seen in response to stress, but the stress effect was not seen in adrenalectomized rats. The effects of administration of tryptophan and cortisone were additive, and the administration of Actinomycin D eliminated the response to cortisone but did not affect the response to tryptophan. The action of tryptophan is apparently one of substrate induction, similar to what we have already seen; in other words, it works primarily by increasing substrate availability to already present enzyme, thus preventing enzyme degradation. Cortisone activity, however, is *dependent on* the production of new RNA from the DNA template since its action is inhibited by the action of actinomycin. The function of the corticoids would seem, therefore, to be to promote induction of RNA synthesis. Evidence has been found that other enzyme systems operate in this fashion, for example, tyrosine transaminase, phosphoenolpyruvic carboxykinase. It is entirely possible that the induction of enzyme synthesis through the action of steroid hormones is a type of induction or derepression available to highly differentiated organisms and unavailable to less differentiated microbial organisms (Jacob & Monod, 1961b).

It is an established fact that hormonal regulation of body processes is intimately involved with regulatory processes stemming from the nervous system. Increasingly active interest is being taken in the study of these neuroendocrine mechanisms. Although our knowledge of the relations between the hypothalamus and the pituitary is still extremely hypothetical, it seems likely that the primary site of the production of neurohumoral secretions regulating the production of tropic hormones is in the various nuclei of the hypothalamus, and that the neurosecretions are carried by the hyphyseal portal veins to the adenohypophysis (Scharrer & Scharrer, 1963). This would suggest that the hormonal feedback may have something to do with the regulation of production of tropic hormones by neural cells. There is already a fairly clear connection established between osmotic pressure of the blood and the production of antidiuretic hormone by the cells of the supraoptic nucleus of the hypothalamus. As a further indication of the integral functioning of the neural and hormonal systems, one cannot fail to be impressed by the sensitivity of the response of the pituitary-adrenal axis to stress. The possibility that such sensitive and profound reactions, which affect and involve the functioning of the entire organism, may be mediated by genetic mechanisms at the level of cellular regulation brings the genetic component of the adaptive process into much closer involvement with the moment-to-moment dynamics of the organism's functioning.

It is entirely speculative, but not altogether impossible, to begin to think of neural processes in terms of the same type of regulatory mechanisms. If it can be established that the neural impulse is mediated by chemical mediator substances such as adrenaline or acetylcholine, it is entirely reasonable to assume that the production of such substances is controlled through enzyme systems of one kind or other. The evidence provided by the electron microscope that the nerve cell possesses an endoplasmic reticulum and a protein synthesizing mechanism (Palay & Palade, 1955) and the further evidence supplied by Weiss & Cavanaugh (1959) that the neuron is continually producing cellular materials suggests that the nerve cell might also function through the control of protein synthesis by genetic mechanisms. The further study of enzymes involved in the acetylcholine system in the generation and transmission of nerve impulses will undoubtedly show the integral relationship between the activation of such enzyme systems and the functioning of the nerve cell (Nachmansohn, 1961, 1962).

Interest has also been developing in the biochemical aspects of learning. The relation between acetylcholinesterase concentrations and previous learning has been approached experimentally (Krech, Rosenzweig, & Bennett, 1960; Rosenzweig, Krech, & Bennett, 1960; Thomas, 1962) and acetylcholinesterase levels have also been found to be related both to total quantity of stimulation and to genetic differences in maze learning ability in rats (Smith, 1962). Such changes in enzyme levels in relation to neural cell activation are presumably under genetic control. It has also become apparent that RNA levels are elevated in active cells and Hyden (Hyden & Egyhazi, 1962) has even been able to demonstrate alterations in the specific nuceotide composition of RNA after learning. It is not impossible that the functional activity of the neuron, even in higher cerebral functions like learning, brings about structural modifications of the participating neurons which involve cellular regulatory mechanisms at the genetic level.

The implication of cellular regulatory mechanisms for the understanding of more complex forms of behavior must remain for the present highly speculative. Yet the few pieces of the picture which can be substantiated by available evidence suggest that the progress of our knowledge is forcing on us new ways of thinking about the complexities of the response of the organism. The study of genetic regulatory mechanisms brings the genetic apparatus into much clearer focus as an intimately reactive component of the total ongoing and dynamic adaptation of the organism. The dichotomy between genetically determined structure and environmentally determined function can no longer be permitted to dominate the thinking about the behavioral aspects of genetics. Structure and function can no longer be dissociated in our thinking, even at the most basic levels of molecular organization. It is apparent, futhermore, that the most basic levels of molecular transcription are intimately connected with, and reciprocally interact with, the more complex levels of organization *and* function which characterize the organism's life history.

From this point of view, it is interesting to speculate that the development of our thinking about genetic regulatory mechanisms may undercut the traditional arguments about the etiology of schizophrenia. Recent reviews of front-running biochemical candidates (Kety, 1959) suggest a wide variety of metabolic factors which might be involved in schizophrenic etiology. The structural similarities between mescaline and epinephrine suggests a defect in the methylation of epinephrine as a possibility. Gottlieb and his group (Beckett, Senf, Frohman, Tourney, & Gottlieb, 1962) have been able to relate primary symptoms to a defect in the moblization after stress of fructose-1, 6-diphosphate, an important intermediate in the metabolism of carbohydrates. They believe that this is related to a failure in the mechanism for the production of ATP, the universal energizing intermediate for biological activation. Other abnormalities in the metabolism of the aminoacid tryptophan have been established (Price, Brown, & Peters, 1959). The list can readily be extended, but the significant point is that it is generally agreed that the schizophrenic process involves some sort of metabolic disturbance. Moreover, it can be pointed out that such metabolic abnormalities are mediated by alterations in the enzyme systems controlling metabolite production. Such enzyme systems are regulated by genetic mechanisms. Consequently, the abnormal metabolic processes which accompany the schizophrenic syndrome may be effected either by alterations in the structural genes or by activation or inhibition of regulator genes. In the latter instance, it is no longer possible to regard genetic factors in isolation from environmental factors which can influence genetic regulation by way of induction or repression. It is possible in these terms to think of a sequence of events deriving from the environmental interaction or life history of a schizophrenic subject and terminating in the modification or modulation of the genetic apparatus, which would account for metabolic alterations and give rise to the symptom complex characteristic of the disease. It would seem, then, that the isolation and validation of genetic influences in populations must be subject to a further refinement, namely, that the frequencies of abnormalities be shown to derive from the alteration of structural genes rather than from the activation or inhibition of regulator genes.

The mechanisms and specific interactions involved in the genesis of mental illness remain to be determined. But progress in science is frequently purchased at the cost of new refinements in our categories of thinking. The understanding of the regulatory function of the genetic mechanism is perhaps one such refinement which can contribute to progress in our thinking about the interaction of genetic and environmental factors in the etiology of mental illness and in the adaptation of the human organism.

REFERENCES

Ames, B. N., & Garry, B. Coordinate repression of the synthesis of four histidine biosynthetic enzymes by histidine. *Proceedings of the National*

Academy of Sciences of the United States of America, 1959, **45**, 1453–1461.

Avery, O. T. A., MacLeod, C. M., & McCarthy, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *Journal of Experimental Medicine*, 1944, **79**, 137-157.

Beadle, G. W. The new genetics: The threads of life. In, *Britannica book of the year, 1964*. Chicago: Encyclopedia Britannica, 1964. Pp. 44–72.

Beckett, P. G. S., Senf, R., Frohman, C. E., Tourney, G., & Gottlied, J. S. Relations between energy transfer systems and the symptoms of schizophrenia. *American Journal of Psychiatry*, 1962, **118**, 995–1003.

Brenner, S., Jacob, F., & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, 1961, **190**, 576–581.

Changeux, J. P. The feedback control mechanism of biosynthetic L-threonine deaminase by L-isoleucine. *Cold Spring Harbor Symposia on Quantitative Biology*, 1961, **26**, 313–318.

Cocito, C., & Vogel, H. J. Heritable lowering of an enzyme level and enzyme repressibility observed upon continuous cultivation of Escherichia coli in the presence of a repressor. In, *Proceedings of the Tenth International Congress of Genetics*, Vol. 2, Toronto: Univer. Toronto Press, 1958. P. 55.

Crick, F. H. C., Barnett, L., Brenner, S., & Watts-Tobin, R. J. General nature of the genetic code for proteins. *Nature*, 1961, **192**, 1227–1232.

Gale, E. F., & Folkes, J. P. The assimilation of amino-acids by bacteria. *Biochemical Journal*, 1955, **59**, 661–684.

Garrod, A. E. *Inborn errors of metabolism*. London: Frowde, Hodder, & Stoughton, 1923.

Gerhart, J. C., & Pardee, A. B. The enzymology of control by feedback inhibition. *Journal of Biological Chemistry*, 1962, **237**, 891–896.

Griesbach, W. E., & Purves, H. D. Studies of experimental goitre: V. Pituitary function in relation to goitrogenesis and thyroidectomy. *British Journal of Experimental Pathology*, 1943, **24**, 174–184.

Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. F., & Zamecnik, P. C. A soluble ribonucleic acid intermediate in protein synthesis. *Journal of Biological Chemistry*, 1958, **231**, 241-257.

Hogness, D. S., Cohen, M., & Monod, J. Studies on the induced synthesis of β-galactosidase in Escherichia coli: The kinetics and mechanism of sulfur incorporation. *Biochimica et Biophysica Acta*, 1955, **16**, 99–116.

Hyden, H., & Egyhazi, E. Nuclear RNA changes of nerve cells during a learning experiment in rats. *Proceedings of the National Academy of Sciences of the United States of America*, 1962, **48**, 1366–1373.

Jacob, F. Genetic control of viral functioning. In, *The Harvey lectures*. New York: Academic Press, 1960. Pp. 1–39.

Jacob, F., & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 1961, **3**, 318–356. (a)

Jacob, F., & Monod, J. On the regulation of gene activity. *Cold Spring Harbor Symposia on Quantitative Biology*, 1961, **26**, 193. (b)

Jacob, F., & Monod, J. Genetic repression, allosteric inhibition and cellular differentiation. In M. Locke (Ed.), *Cytodifferentiation and macromolecular synthesis*. New York: Academic Press, 1963. Pp. 30–64.

Jacob, F., Perrin, D., Sanchez, C., & Monod, J. L'opéron: Groupe de gènes à expression coordonée par un opérateur. *Compte Rendu de Academie des Sciences, Paris*, 1960, **250**, 1727–1729.

Jones, O. W., Jr., & Nirenberg, M. W. Qualitative survey of RNA codewords. *Proceedings of the National Academy of Sciences of the United States of America*, 1962, **48**, 2115–2123.

Karlson, P. New concepts on the mode of action of hormones. *Journal of Endocrinology*, 1962, **24**, iv.

Kety, S. S. Biochemical theories of schizophrenia. *Science*, 1959, **129**, 1528–1532, 1590–1596.

Knox, W. E., & Auerbach, V. H. The hormonal control of tryptophan peroxidase in the rat. *Journal of Biological Chemistry*, 1955, **214**, 307–313.

Knox, W. E., Auerbach, V. H., & Lin, E. C. C. Enzymatic and metabolic adaptations in animals. *Physiological Reviews*, 1956, **36**, 164–254.

Kramer, M., & Straub, F. B. Role of specific nucleic acid in induced enzyme synthesis. *Biochimica et Biophysica Acta*, 1956, **21**, 401–403.

Krech, D., Rosenzweig, M. R., & Bennett, E. L. Effects of environmental complexity and training on brain chemistry. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 509–519.

Lark, K. G., & Lark, C. Changes during the division cycle in bacterial cell wall synthesis, volume, and ability to concentrate free amino-acids. *Biochimica et Biophysica Acta*, 1960, **43**, 520–530.

Matovinovic, J., & Vickery, A. L. Relation of nucleic acids to the structure and function of the guinea pig thyroid gland. *Endocrinology*, 1959, **64**, 149–159.

McFall, E., Pardee, A. B., & Stent, G. S. Effects of radiophosphorus decay on some synthetic capacities of bacteria. *Biochimica et Biophysica Acta*, 1958, **27**, 282–297.

Monod, J., Changeux, J. P., & Jacob, F. Allosteric proteins and cellular control systems. *Journal of Molecular Biology*, 1963, **6**, 306–329.

Nachmansohn, D. Chemical factors controlling nerve activity. *Science*, 1961, **134**, 1962–1968.

Nachmansohn, D. Basic aspects of nerve activity explained by biochemical analysis. *Journal of the American Medical Association*, 1962, **179**, 639–643.

Neidhardt, F. C., & Fraenkel, D. G. Metabolic regulation of RNA synthesis in bacteria. *Cold Spring Harbor Symposia on Quantitative Biology*, 1961, **26**, 63–74.

Nirenberg, M. W., & Mathaei, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America,* 1961, **47**, 1588–1602.

Palay, S. L., & Palade, G. E. The fine structure of neurons. *Journal of Biophysical and Biochemical Cytology,* 1955, **1**, 69–88.

Pardee, A. B., Jacob, F., & Monod, J. Sur l'expression et le rôle des allèles "inductible" et "constitutif" dans la synthese de la β-galactosidase chez des zygotes d'Escherichia coli. *Compte Rendu de Academie des Sciences, Paris,* 1958, **246**, 3125–3128.

Pardee, A. B., Jacob, F., & Monod, J. The genetic control and cytoplasmic expression of "inductibility" in the synthesis of β-galactosidase by E. coli. *Journal of Molecular Biology,* 1959, **1**, 165–178.

Price, J. M., Brown, R. R., & Peters, H. A. Tryptophan metabolism in porphyria, schizophrenia and a variety of neurologic and psychiatric diseases. *Neurology,* 1959, **9**, 456.

Rogers, P., & Novelli, G. D. Formation of ornithine transcarbamylase in cells and protoplasts of Escherichia coli. *Biochimica et Biophysica Acta,* 1959, **33**, 423–436.

Rosenzweig, M. R., Krech, D., & Bennett, E. L. A search for relations between brain chemistry and behavior. *Psychological Bulletin,* 1960, **57**, 476–492.

Rotman, B., & Spiegelman, S. On the origin of the carbon in the induced synthesis of β-galactosidase in Escherichia coli. *Journal of Bacteriology,* 1954, **68**, 419–429.

Schaechter, M. Patterns of cellular control during unbalanced growth. *Cold Spring Harbor Symposia on Quantitative Biology,* 1961, **26**, 53–62.

Scharrer, E., & Scharrer, B. *Neuroendocrinology.* New York: Columbia Univer. Press, 1963.

Shepherdson, M., & Pardee, A. B. Preparation and crystallization of aspartate transcarbamylase of Escherichia coli. *Bacteriological Proceedings,* 1960, **60**, 172. (a)

Shepherdson, M., & Pardee, A. B. Production and crystallization of aspartate transcarbamylase. *Journal of Biological Chemistry,* 1960, **235**, 3233–3237. (b)

Smith, C. E. Is memory a matter of enzyme induction? *Science,* 1962, **138**, 889–890.

Smythies, J. R. *Schizophrenia: Chemistry, metabolism and treatment.* Springfield, Ill.: Thomas, 1963.

Spiegelman, S., Halvoørsen, H. O., & Ben-Ishai, R. Free amino-acids and the enzyme-forming mechanism. In W. D. McElroy, & H. B. Glass (Eds.), *Symposium on amino-acid metabolism.* Baltimore: Johns Hopkins Press, 1955. Pp. 124–170.

Stevens, C. E., D'Angelo, S. A., Paschkis, K. E., Canterow, A., & Sunderman, F. W. The response of pituitary-thyroid system of the guinea pig to low environmnetal temperature. *Endocrinology,* 1955, **56**, 143–156.

Thomas, G. J. Neurophysiology of learning. In, *Annual Review of Psychology,* Stanford, Calif.: Annual Reviews, 1962. Pp. 71–106.

Torriani, A. Influence of inorganic phosphate in the formation of the phosphatases by E. coli. *Biochimica et Biophysica Acta,* 1960, **38**, 460–468.

Umbarger, H. E. Feedback control by endproduct inhibition. *Cold Spring Harbor Symposia on Quantitative Biology,* 1961, **26**, 301–312.

Watson, J. D., & Crick, F. H. C. Genetical implications of the structure of deoxyribonucleic acid. *Nature,* 1953, **171**, 964–967.

Weiss, P., & Cavanaugh, M. W. Further evidence of perpetual growth of nerve fibers: Recovery of fiber diameter after the release of prolonged constrictions. *Journal of Experimental Zoology,* 1959, **142**, 461–473.

Yates, R. A., & Pardee, A. B. Control by uracil of formation of enzymes required for orotate synthesis. *Journal of Biological Chemistry,* 1957, **227**, 677–692.

Zalokar, M. Ribonucleic acid and the control of cellular processes. In D. M. Bonner (Ed.), *Control mechanisms in cellular processes.* New York: Ronald Press, 1961. Pp. 87–140.

(Received March 26, 1965)

Psychological Bulletin
1965, Vol. 64, No. 3, 217-224

LATERALITY OF SUBCORTICAL INVOLVEMENT AND PSYCHOLOGICAL FUNCTIONS[1]

MANUEL RIKLAN AND ERIC LEVITA[2]

St. Barnabas Hospital, New York

Differences in psychological effects between lateralized cortical and subcortical brain lesions were assessed within a neuropsychological framework. The relative predominance of specific and nonspecific systems in the elaboration of incoming data at these cerebral levels and their participation in cognition and perceptual events was described. It was concluded that bilateral primacy of subcortical diffuse systems in contradistinction to unilateral primacy of cortical specific systems may account for observed differences between lateralized cortical and subcortical brain lesions.

A recent multidisciplinary symposium on interhemispheric relations and cerebral dominance (Mountcastle, 1962), monographs on language disorders (Osgood & Myron, 1963; Zangwill, 1960), and experimental studies on the role of the corpus callosum and other interhemispheric commissures (Sperry, 1961; Myers, 1961), attest to a renewed interest in hemispheric differences and behavior. Since Broca's (1861) pronouncements relating speech functions to the left hemisphere, the pervading belief has been that this hemisphere is "dominant," at least for language activities. In the past several years, more intensive laboratory and clinical investigations have been undertaken to delineate other facets of hemispheric dominance and language, and to explore such areas of behavior as perception and cognition in their relationship to functional differences between the brain hemispheres. It is the purpose of this paper to review and reintegrate existing concepts of hemispheric role in humans, particularly as revealed by the effects of surgically induced unilateral basal ganglia or thalamic lesions on psychological functions.

CURRENT CONCEPTS OF CORTICAL LATERALITY

Much current thinking on both laterality and hemispheric dominance continues to emphasize a language-nonlanguage bifurcation, tends to be clinical in nature, and, in many instances, utilizes the convenient Wechsler-Bellevue Intelligence Scale (WBIS). In an early study, Anderson (1951) administered this scale to a group of individuals with brain-damaged dominant hemispheres and an equivalent group of "nondominant" patients and reported a significantly greater loss on verbal tests for the dominant hemisphere group and a correspondingly greater loss on performance tasks for the nondominant population. A number of confirmatory studies have since appeared (Bauer & Becka, 1954; Costa & Vaughan, 1962; Heilbrun, 1956; McFie, 1960; Morrow & Mark, 1955; Reitan, 1955; Stark, 1961).

The mainstream of present-day beliefs concerning the effect of brain lesions on language is exemplified by Bauer and Wepman (1955) and Boone (1959), who agree that cerebral dominance seems unique to the left hemisphere. However, Heilbrun (1956), in assessing the effects of right hemisphere lesions, failed to find differences in spatial performance. This result is compatible with two other studies (Battersby, Bender, Pollack, & Kahn, 1956; Piercy & Smyth, 1962), in which spatial representation was found to be bilateral but unequal, with differences in degree rather than kind. For this function, the evidence is very much against strict unilateral representation (Osgood & Myron, p. 53).

In the realm of cerebral dominance and language, a similar changing emphasis is also apparent. Instances of spontaneous recovery of language following brain injury indicate that there must be some replication of function "either between the two hemispheres or between regions of the dominant hemisphere [Osgood & Myron, 1963, p. 49]." Cerebral dominance may in all probability be a graded characteristic, according to Zangwill (1960, p. 27), Eisenson (1962), and Critchley (1962, pp. 208–214), the last two of whom stressed the importance of the right cerebral hemisphere in language formulation. Critchley conceived of speech as a spectrum ranging from the "normal" subject, at one extreme, through the linguistic pattern of nondominant hemisphere defect to a full-fledged aphasia from disease of the dominant half of the brain.

---

[1] From the Psychological Services of St. Barnabas Hospital, New York City. Supported in part by Grant No. RD-1274-M from the Vocational Rehabilitation Administration, Washington, D. C.

[2] Thanks are expressed to Irving S. Cooper, director, Department of Neurosurgery, for his encouragement and continuous support.

Chapman and Wolff (1959) extended this formulation to the "highest integrative functions" which they found to be directly related to the mass of cerebral tissue lost regardless of site and side. This corroborates Lashley's (1929) earlier principle of mass action, and Hecaen's (1959) concept of equipotentiality, implying that either hemisphere can assume the functions of the other, although at a reduced level. In essence, the potential of the "minor" hemisphere may not ordinarily be utilized maximally except in the case of a lesion to the major hemisphere. Analogous experimental work involving corpus callosum sectioning (Sperry, 1961; Myers, 1961) has demonstrated the independence of such surgically separated hemispheres with respect to learning and memory, since separate memory trace systems seem to be developed in both hemispheres following learning primarily involving one hemisphere (Ebner & Myers, 1962a, 1962b). It should be noted finally that a report of Doehring and Reitan (1962) illustrates the need for differentiating levels of behavior in properly assessing the psychological effects of lateralized brain lesions. Reviews by Meyer (1961), Reitan (1962), and most recently Piercy (1964) also emphasize a lack of clear-cut distinctions, despite some apparent specialization in lateralized functions. Contemporary psychological concepts of functional laterality in humans have been derived predominantly from studies of patients with insults on their cerebral cortices, including epileptics, hemiplegics, brain-operated or traumatically injured. In many reported cases, obviously subcortical lesions were present concomitant with cortical damage. Not all previously described investigators have taken identical viewpoints, and proposals have been made for a theory of graded differentials. Recent evidence, moreover, suggests a possible unwarranted equation of differences in degree with differences in kind, concurrent with a failure to account for the role of subcortical structures in psychological behavior.

PSYCHOLOGICAL EFFECTS OF SUBCORTICAL BRAIN SURGERY

Psychological investigations were undertaken with individuals undergoing unilateral subcortical surgery, chemothalamectomy and chemopallidectomy (Cooper, 1956, 1961) or cryothalamectomy (Cooper, 1964), for the relief of tremor and rigidity in Parkinson's disease. Such operations involve surgical production of a lesion in the mesial globus pallidus (in chemopallidectomy) or the ventrolateral nucleus of the thalamus (in chemothalamectomy and cryothalamectomy). Neuroanatomical studies have confirmed the site of lesion placement (Caracalos, Levita, & Cooper, 1962; Cooper, Bergmann, & Caracalos, 1963; Dierssen, Bergmann, Gioino, & Cooper, 1962). Both of these structures eventually project largely to the ipsilateral cortex.

In an initial study by Riklan, Diller, Weiner, and Cooper (1960), using the WBIS, both subcortical hemisphere groups demonstrated significant decreases in verbal scores, while only right-brain operatees showed a like decrease in performance tests within three weeks following surgery. No significant differences in postoperative intellectual effects were found between lesion sites. A return to preoperative status for the total group was observed on follow-up assessment, nine months postoperatively. These results appeared compatible with Sheer's (1956) investigation involving surgically produced cortical lesions in which immediate postoperative decline was reversed in follow-up testing. Moreover, an association between left hemisphere subcortical lesions and deficits in language seemed to be verified. It was later found (Riklan & Diller, 1961) that immediately following surgery, patients who had had an operation on the right basal ganglia regressed in visual-motor performance, while the left-hemisphere group evidenced no like change. However, results of the Human Figure Drawing Test administered before and after chemosurgery (Riklan, Zahn, & Diller, 1962) suggested that body-image perception underwent postoperative modification irrespective of side of subcortical surgery. It was noted then that both hemispheres may contribute to appropriate body image integration. This finding questioned earlier studies relating body-image function primarily to the right hemisphere (Critchley, 1953; Kolb, 1959).

Further exploration into the nature of presumed lateralized differences and thalamic lesions was achieved through a factor analysis of 35 variables from the WBIS, the Bender-Gestalt Test, Human Figure Drawing Test, and the Rorschach Psychodiagnostic Test (Riklan & Levita, 1964). Observed lateralized differences were primarily in factor loadings, with greater verbal loadings for the left-brain operatees and higher nonverbal loadings for the right-brain group. However, the apparent qualitative similarity of psychological change factors, as shown in Table 1, led to the hypothesis that certain psychological functions might be subserved by systems available to both subcortical regions. These findings opened to further consideration earlier reports which either failed to weigh subcortical laterality (Freed & Pastor, 1951; Gillingham, 1961; Krayenbuehl & Yasargil, 1961; Orchinik, 1960; Orchinik, Koch, Wycis, Freed, & Spiegel, 1949), or ascribed more definitive differences in functions to the left and right subcortical structures (McFie, 1960).

TABLE 1

SUMMARY OF PRE–POSTOPERATIVE CHANGE FACTORS

($N = 71$)

Change factors	Left-brain variance (%)	Right-brain variance (%)
Verbally mediated cognition	19	9
Nonverbally mediated cognition	13	19
Visual-motor construction expansion	12	12
Perceptual approach	11	12
Body image integration	11	12
Ideational productivity	9	9
Impulse control	9	—
Emotional reactivity	6	—
Perceptual integration	—	9
Factors not identified	10%	18%

Note.—From Riklan and Levita (1964).

Verbally and nonverbally mediated cognitive tasks were used for further elaboration and clarification of the nature of subcortical

TABLE 2

VARIMAX ROTATION OF CENTROID FACTORS FOR LEFT AND RIGHT SUBCORTICAL INVOLVEMENT

($N = 24$)

Tests	Factor						H^2	
	1R	1L	2R	2L	3R	3L	R	L
Raven's matrices	.86	.66	.11	.20	−.28	.13	.83	.49
Block design	.71	.71	.19	.26	.11	−.15	.55	.59
Vocabulary	.09	.18	.78	.74	−.02	.19	.62	.62
Proverbs	.13	.13	.69	.77	−.21	.24	.54	.67
Digits	.48	.43	.58	.18	.13	.58	.59	.56
Mental control	.32	−.07	.43	.02	.38	.59	.43	.36
Minnesota form board	.26	.60	.38	−.23	−.58	.08	.55	.42
Information	.02	−.05	.01	.40	.41	−.19	.16	.20
Common variance	.39	.39	.42	.39	.19	.22	1.00	1.00

Note.—Adapted from Levita and Riklan (1965).

laterality. No significant behavioral differences were obtained as a function of lateralized involvement, based on groups of predominantly right-handed subjects (92% right-handed). Particularly, no differences were found in unoperated parkinsonians with predominant left- or right-sided neurological symptoms, that is, presumably right- or left-sided basal ganglia damage (Levita, Riklan, & Cooper, 1964). Moreover, in a partial cross validation and extension of Riklan and Levita's (1964) original factorial study, a battery of cognitive-perceptual tests was administered to a group of unoperated parkinsonians and also subjected to factor analysis (Levita & Riklan, 1965). Two major, almost identical factors—specifically a visual-spatial and verbal discrimination factor for each hemisphere group—were found (shown in Table 2). In the case of surgery, no substantial immediate postoperative differences were found between left and right thalamic lesions as a function of symbolization, whether verbally or nonverbally mediated (Levita, Riklan, & Cooper, 1964). Finally, laterality and site of surgery were not pertinent to verbal and perceptual changes.

In the light of the foregoing findings, the conclusion seems warranted that laterality of predominantly subcortical impairment does not result in the kind of differences reported for cortical lesions with or without some subcortical involvement. It follows that reassessment of lateralized aspects of neural organization as it pertains to integrative behavior is indicated.

Discussion

Physiological Considerations

Recent electrophysiological studies have shown bilateral hemispheric changes following unilateral subcortical surgery. In particular, electroencephalographic changes immediately following such surgery have frequently shown bilateral decrease in speed of rhythm and increase in voltage (Bravo, 1956; Lin, 1964). The bilateral influence of the basal ganglia was corroborated by pathophysiological studies on the role of ipsilateral lesions in hemichorea and hemiballismus (Dierssen, Gioino, & Cooper, 1961) and the finding of Penfield and Roberts (1959, pp. 207–208) that a bilaterally distributed subcortical zone is responsible for the integration of speech.

Complementary studies involving thalamic stimulation during unilateral stereotactic surgery also confirmed a contralateral as well as an ipsilateral component for this region. Both Ervin and Mark (1960; 1964) and Walter, Rand, Crandall, Markham, and Adey (1963) observed bilateral, cortical-evoked responses to stimulation of specific as well as nonspecific thalamic nuclei. It is thus conceivable that specific systems contribute mainly, though not exclusively, to ipsilateral cortical responses while nonspecific systems mediate primarily bilateral reactions, although the manner of interhemispheric transmission is not clear.

While the ventrolateral nucleus of the thalamus and the mesial globus pallidus sites of the surgical lesion project mainly to the unilateral cortex, the physiological disruption itself may be bilateral in nature, involving, (*a*) the interaction of the specific lesion with nonspecific subcortical systems and (*b*) the cortex, bilaterally.

Laboratory studies have yielded analogous evidence and stressed the importance of nonspecific systems in neural functions. The alerting or arousal of the central nervous system has been attributed to the brain-stem reticular activating system (Magoun, 1954, 1963; Moruzzi & Magoun, 1949). A functionally related system, the nonspecific thalamo-cortical projections, also modulates sensory and motor functions of the cortex, according to Morison and Dempsey (1942a, 1942b); Dempsey and Morison (1942a, 1942b), and Jasper (1949; 1960). It is of special relevance that the bilateral influence of the diffuse thalamo-cortical system has been confirmed also in laboratory experiments (Albe-Fessard & Rouguel, 1958; Enomoto, 1959). This bilaterality is not due to cortico-cortical connections, transfer of impulses through the corpus callosum, or the anterior or posterior commissures (Enomoto, 1959; Jasper, 1960). More recently a bilateral, basal ganglia influence upon ascending nonspecific activity was reported (Krauthamer & Albe-Fessard, 1965).

Behavioral Considerations

Research on cerebral activating systems has served to focus interest on the alerting or activating aspects of behavior. According to a review by Samuels (1959), constructs such as attention, perception, drive, and positive and negative reinforcement may possess a common nonspecific activating component in addition to their specific sensorimotor properties. Lindsley (1951, 1956, 1957, 1960) added emotion and learning to the behavioral categories subserved by neural activating processes. It would therefore appear that a more complete description of behavior requires assessment of activating as well as directional aspects (Duffy, 1934, 1957, 1962; Malmo, 1959). A particularly striking example of the failure to differentiate between activation and direction is illustrated by emotion, where both are included in a single concept. Other areas of behavior may similarly be analyzed.

The importance of activation or arousal in perceptual functions was demonstrated in animal studies in which *moderate* electrical stimulation of diffuse systems facilitated perceptual sensitivity and discrimination (Fuster, 1958; Lindsley, 1958). Moreover, the perceptual and cognitive impairments seen in sensory deprivation have been attributed by Lindsley (1961) to altered sensory input affecting the ordinary operation of nonspecific sensory systems. With respect to cognition, the "orienting reflex" has been considered to play a preparatory, preadaptive role (Razran, 1961; Sokolov, 1960, 1963). This reflex represents an alerting reaction to novel stimuli which may constitute "the material basis" from which cognition emerges (Razran, 1961, p. 119) and may also be considered equivalent to the behaviorists' "operant level" (Keller & Schoenfield, 1950, p. 76). This energizing response has been related to action of the reticular formation.

Optimal alerting would seem to facilitate an organism's perceptual as well as cognitive adaptability. In this context, perceptual readiness is described by Bruner (1957) as minimizing the surprise of an organism and limiting its need for searching. Moreover, for Bruner, perception involves an act of categorization. Such categorization or discrimination embraces both perceptual and cognitive functions, as exemplified by responses to psychological tests. The authors therefore propose that such responses incorporate specific and nonspecific alerting in combination.

A Neuropsychological Synthesis

It was suggested earlier that subcortical impairment does not result in the kind of lateralized effects seen in primarily cortical lesions, the differences, where found, being in degree rather than in kind. For both physiological and psychological events, specific and diffuse processes operate concomitantly. Concomitance of physiological and psychological events is not meant to imply an identity or semantic confusion between these two subuniverses of discourse.

At the thalamic level, diffuse and specific systems interact in the elaboration of information. Diffuse bilateral processes contribute alerting or attention components while concurrent specific processes contribute more definitive information. At this level, the less completely organized data may require greater diffuse activity to achieve elaboration. Within the cortical mantle, when further integration has occurred, less diffuse driving may be required. At this level, too, lateral differentiation has occurred within a framework of more definitive integration. In short, for sensorimotor integration, relative predominance of bilateral diffuse systems is assumed at the subcortical level, whereas at the cortical level more specific unilaterally distributed influences may predominate.

Surgically induced subcortical lesions, affecting data processing at a predominantly diffuse stage, might be expected to disrupt higher order behaviors regardless of lesion laterality. This proposition has been sufficiently confirmed by the series of subcortical studies previously enumerated. On the other hand, results of the cortical studies have generally shown a greater degree of lateralization. Such results are also compatible with our concept concerning greater differentiation and lateralization at cortical levels.

The foregoing analysis has emphasized the role of incoming data involving both specific and diffuse processes. Yet the nature of cognition and perception is such that outgoing and modulating events also participate. Cognitive and perceptual behaviors involve all levels of sensorimotor integration, and a breakdown or lesion at any level would affect the entire system and therefore the ultimate response.

REFERENCES

Albe-Fessard, D., & Rougueul, A. Activités d'origine somesthesique evoquées sur le cortex nonspecifique du chat anesthesie au chloralose: Role du centre median du thalamus. *Electroencephalography and Clinical Neurophysiology,* 1958, **10**, 131–152.

Anderson, A. L. The effect of laterality localization of focal brain lesions on the Wechsler-Bellevue subtests. *Journal of Clinical Psychology,* 1951, **7**, 149–153.

Battersby, W. S., Bender, M. B., Pollack, M., & Kahn, R. L. Unilateral "spatial agnosia" ("inattention") in patients with cerebral lesions. *Brain,* 1956, **79**, 68–93.

Bauer, R. W., & Becka, D. M. Intellect after cerebral vascular accident. *Journal of Nervous and Mental Disease,* 1954, **120**, 379–384.

Bauer, R. W., & Wepman, J. M. Lateralization of cerebral functions. *Journal of Speech and Hearing Disorders,* 1955, **20**, 171–177.

Boone, D. R. Communication skills and intelligence in right and left hemiplegics. *Journal of Speech and Hearing Disorders,* 1959, **24**, 241–248.

Bravo, G., & Cooper, I. S. Early electroencephalographic observations following chemopallidectomy. *Journal of the American Geriatrics Society,* 1956, **4**, 107–111.

Broca, P. Remarques sur le siège de la faculté du language articulé. *Bulletin de la Societé d'anatomique de Paris,* 1861, **6**, 336–357.

Bruner, J. S. On perceptual readiness. *Psychological Review,* 1957, **64**, 123–152.

Caracalos, A., Levita, E., & Cooper, I. S. A study of roentgeno-anatomic lesion location and results in cryosurgery of the basal ganglia. *St. Barnabas Hospital Medical Bulletin,* 1962, **1**, 24–31.

Chapman, L. F., & Wolff, H. G. The cerebral hemispheres and the highest integrative functions of man. *American Medical Association Archives of Neurology,* 1959, **1**, 357–424.

Cooper, I. S. *The neurosurgical alleviation of parkinsonism.* Springfield, Ill.: Charles C Thomas, 1956.

Cooper, I. S. *Parkinsonism: Its medical and surgical therapy.* Springfield, Ill.: Charles C Thomas, 1961.

Cooper, I. S. Cryogenic surgery in the geriatric patient. *Journal of the American Geriatrics Society,* 1964, **12**, 813–855.

Cooper, I. S., Bergmann, L., & Caracalos, A. Anatomic verification of the lesion which abolishes parkinsonian tremor and rigidity. *Neurology,* 1963, **13**, 779–787.

Costa, L. D., & Vaughan, H. G. Performance of patients with lateralized cerebral lesions. I. Verbal and perceptual tests. *Journal of Nervous and Mental Disease,* 1962, **134**, 162–168.

Critchley, M. *The parietal lobes.* London: Arnold, 1953.

Critchley, M. Speech and speech loss in relation to the quality of the brain. In V. B. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance.* Baltimore: Johns Hopkins Press, 1962. Pp. 208–214.

Dempsey, E. W., & Morison, R. S. The interaction of certain spontaneous and induced cortical potentials. *American Journal of Physiology,* 1942, **135**, 301–308. (a)

Dempsey, E. W., & Morison, R. S. The production of rhythmically recurrent cortical potentials after localized thalamic stimulation. *American Journal of Physiology,* 1942, **135**, 293–300. (b)

Dierssen, G., Bergmann, L., Gioino, G., & Cooper, I. S. Surgical lesions affecting parkinson symptomatology. *Acta neurochirurgica,* 1962, **10**, 125–133.

Dierssen, G., Gioino, G., & Cooper, I. S. Participation of ipsilateral hemisphere lesions in the pathology of hemichorea and hemiballismus. *Neurology,* 1961, **11**, 894–898.

Doehring, D. G., & Reitan, R. M. Concept attainment of human adults with lateralized cerebral lesions. *Perceptual and Motor Skills,* 1962, **14**, 27–33.

Duffy, E. Emotion: An example of the need for reorientation in psychology. *Psychological Review,* 1934, **41**, 184–198.

Duffy, E. The psychological significance of the concept of "arousal" or "activation." *Psychological Review,* 1957, **64**, 265–275.

Duffy, E. *Activation and behavior.* New York: Wiley, 1962.

Ebner, F. F., & Myers, R. E. Corpus callosum and the interhemispheric transmission of tactual learning. *Journal of Neurophysiology,* 1962, **25**, 380–391. (a)

Ebner, F. F., & Myers, R. E. Direct and transcallosal induction of touch memories in the monkey. *Science,* 1962, **138**, 51–52. (b)

Enomoto, T. F. Unilateral activation of the nonspecific thalamic system and bilateral cortical responses. *Electroencephalography and Clinical Neurophysiology,* 1959, **11**, 219–232.

Eisenson, J. Language and intellectual modifications associated with right cerebral damage. *Language and Speech,* 1962, **5**, 47–53.

Ervin, R. E., & Mark, V. H. Stereotactic thalamotomy in the human. Part II. Physiologic observations on the human thalamus. *Archives of Neurology,* 1960, **3**, 368–380.

Ervin, R. E., & Mark, V. H. Studies of the human thalamus. IV. Evoked responses, in sensory evoked response in Man. *Annals of the New York Academy of Sciences,* 1964, **112**, 81–92.

Freed, H., & Pastor, J. T. Evaluation of the "Draw-A-Person Test" (modified) in thalamotomy with particular reference to the body image. *Journal of Nervous and Mental Disease,* 1951, **114**, 106–120.

Fuster, J. M. Effects of stimulation of brain stem on tachistoscopic perception. *Science*, 1958, **127**, 150.

Gillingham, J. Surgical treatment of parkinsonism. *Transactions of the Medical Society of London*, 1961, **77**, 52–56.

Hecaen, H. Dominance hemisphérique et préférence manuelle. *Évolution Psychiatrique*, 1959, **1**, 1–50.

Heilbrun, A. B. Psychological test performance as a function of lateral localization of cerebral lesion. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 10–14.

Jasper, H. Diffuse projection systems: The integrative action of the thalamic reticular system. *Electroencephalography and Clinical Neurophysiology*, 1949, **1**, 405–420.

Jasper, H. Unspecific thalamo-cortical relations. In J. Field (Ed.), *Handbook of physiology*. Vol. 2. Washington, D. C.: American Physiological Society, 1960. Pp. 1307–1322.

Keller, F. S., & Schoenfeld, W. N. *Principles of psychology*. New York: Appleton-Century-Crofts, 1950.

Kolb, L. Disturbances of the body image. In S. Arieti (Ed.), *American Handbook of Psychiatry*. New York: Basic Books, 1959. Pp. 749–769.

Krauthamer, G., & Albe-Fessard, D. Inhibition of non-specific sensory activities following striopallidal and capsular stimulation. *Journal of Neurophysiology*, 1965, **28**, 100–124.

Krayenbuhl, H., & Yasargil, M. G. Ergebnisse der stereotaktischen Operationen beim Parkinsonismus, insbesondere der doppelseitigen Eingriffe. *Deutsche Zeitschrift für Nervenheilkunde*, 1961, **182**, 530–541.

Lashley, K. S. *Brain mechanisms and intelligence*. Chicago: Univer. Chicago Press, 1929.

Levita, E., & Riklan, M. Laterality of subcortical involvement and cognitive performance: A factor analysis. *Perceptual and Motor Skills*, 1965, **20**, 151–157.

Levita, E., Riklan, M., & Cooper, I. S. Verbal and perceptual functions after surgery of subcortical structures. *Perceptual and Motor Skills*, 1964, **18**, 195–202.

Levita, E., Riklan, M., & Cooper, I. S. Cognitive and perceptual performance in parkinsonism as a function of age and neurological impairment. *Journal of Nervous and Mental Disease*, 1964, **139**, 516–520.

Lin, T. H., Matsumoto, K., & Cooper, I. S. Electroencephalographic changes in parkinsonian patients following cryogenic lesions in the ventrolateral nucleus of the thalamus. *Journal of the American College of Surgeons*, 1964, **42**, 281–286.

Lindsley, D. B. Emotions. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 473–516.

Lindsley, D. B. Physiological Psychology. In P. R. Farnsworth (Ed.), *Annual review of psychology*. Stanford, Calif.: Annual Reviews, 1956. Pp. 323–348.

Lindsley, D. B. Psychophysiology and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 44–104.

Lindsley, D. B. The reticular system and perceptual discrimination. In H. H. Jasper (Ed.), *Reticular formation of the brain*. Boston: Little, Brown, 1958. Pp. 513–534.

Lindsley, D. B. Attention, consciousness, sleep and wakefulness. In J. Field (Ed.), *Handbook of physiology*. Vol. 2. Washington, D. C.: American Physiological Society, 1960. Pp. 1553–1594.

Lindsley, D. B. Common factors in sensory deprivation, sensory distinction and sensory overload. In P. Solomon (Ed.), *Sensory deprivation*. Cambridge: Harvard Univer. Press, 1961. Pp. 174–194.

Magoun, H. W. The ascending reticular system and wakefulness. In J. F. Delafresnaye (Ed.), *Brain mechanisms and consciousness*. Oxford: Blackwell, 1954. Pp. 1–20.

Magoun, H. W. *The waking brain*. (2nd ed) Springfield, Ill.: Charles C Thomas, 1963.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychological Review*, 1959, **66**, 367–386.

McFie, J. Psychological effects of stereotaxic operations for the relief of parkinsonian symptoms. *Journal of Mental Science*, 1960, **106**, 1512–1517.

Meyer, V. Psychological effects of brain damage. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. New York: Basic Books, 1961. Pp. 529–565.

Morison, R. S., & Dempsey, E. W. A study of thalamo-cortical relationships. *American Journal of Physiology*, 1942, **135**, 281–292.

Morison, R. S., & Dempsey, E. W. Mechanism of thalamo-cortical augmentation and repetition. *American Journal of Physiology*, 1943, **138**, 297–308.

Morrow, R. S., & Mark, J. C. The correlation of intelligence and neurological findings on twenty-two patients autopsied for brain damage. *Journal of Consulting Psychology*, 1955, **19**, 282–289.

Moruzzi, G., & Magoun, H. W. Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, 1949, **1**, 455–473.

Mountcastle, V. B. (Ed.) *Interhemispheric relations and cerebral dominance*. Baltimore: Johns Hopkins Press, 1962.

Myers, R. E. Corpus callosum and visual gnosias. In A. Fessard (Ed.), *Brain mechanisms and learning*. Oxford: Blackwell, 1961. Pp. 481–506.

Orchinik, C. W. Some psychological aspects of circumscribed lesions of the diencephalon. *Confinia Neurologica*, 1960, **20**, 292–310.

Orchinik, C. W., Koch, R., Wycis, H. T., Freed, H., & Spiegel, E. A. The effect of thalamic lesions on emotional reactivity (Rorschach and behavior studies). *Proceeding of the Association for Research in Nervous and Mental Diseases*, 1950, **29**, 172–207.

OSGOOD, C., & MYRON, M. *Approaches to the study of aphasia.* Urbana: Univer. Illinois Press, 1963.

PENFIELD, W., & ROBERTS, L. *Speech and brain mechanisms.* Princeton: Princeton Univer. Press, 1959.

PIERCY, M. The effects of cerebral lesions on intellectual functions: A review of current research trends. *British Journal of Psychology,* 1964, **110**, 310–352.

PIERCY, M., & SMYTH, V. O. Right hemisphere dominance for certain nonverbal intellectual skills. *Brain,* 1962, **85**, 775–790.

RAZRAN, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological Review,* 1961, **68**, 81–147.

REITAN, R. M. Certain differential effects of left and right cerebral lesions in human adults. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 474–477.

REITAN, R. M. Psychological deficit. *Annual Review of Psychology,* 1962, **13**, 415–444.

RIKLAN, M., & DILLER, L. Visual motor performance before and after chemosurgery of the basal ganglia in parkinsonism. *Journal of Nervous and Mental Disease,* 1961, **132**, 307–313.

RIKLAN, M., DILLER, L., WEINER, H., & COOPER, I. S. Psychological studies on effects of chemosurgery of the basal ganglia in parkinsonism: 1. Intellectual functioning. *American Medical Association Archives of General Psychiatry,* 1960, **2**, 22–31.

RIKLAN, M., & LEVITA, E. Psychological effects of lateralized basal ganglia lesions: A factorial study. *Journal of Nervous and Mental Disease,* 1964, **138**, 233–240.

RIKLAN, M., ZAHN, T., & DILLER, L. Human figure drawings before and after chemosurgery of the basal ganglia in parkinsonism. *Journal of Nervous and Mental Disease* 1962, **6**, 500-506.

SAMUELS, I. Reticular mechanisms and behavior. *Psychological Bulletin,* 1959, **56**, 1–25.

SHEER, D. W. Psychometric studies. In N. Lewis, C. Landis, & H. E. King (Eds.), *Studies in topectomy.* New York: Grune & Stratton, 1956. Pp. 56–74.

SOKOLOV, E. N. Neuronal models and the orienting reflex. In M. A. B. Brazier (Ed.), *CNS and behavior.* New York: Josiah Macy, Jr. Foundation, 1960. Pp. 187–276.

SOKOLOV, E. N. Higher nervous functions: The orienting reflex. *Annual Review of Physiology.* Stanford, Calif.: Annual Reviews, 1963. Pp. 545–580.

SPERRY, R. W. Cerebral organization and behavior. *Science,* 1961, 133, 1749–1757.

SPERRY, R. W. Some general aspects of interhemispheric integration. In V. B. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance.* Baltimore: Johns Hopkins Press, 1962. Pp. 43–50.

STARK, R. An investigation of unilateral cerebral pathology with equated verbal and visual-spatial tasks. *Journal of Abnormal and Social Psychology,* 1961, **62**, 282–287.

WALTER, R. D., RAND, R. W., CRANDALL, P. H., MARKHAM, C. H., & ADEY, W. R. Depth electrode studies of thalamus and basal ganglia. *Archives of Neurology,* 1963, **8**, 388–397.

ZANGWILL, O. L. *Cerebral dominance and its relation to psychological function.* Springfield, Ill.: Charles C Thomas, 1960.

(Received April 8, 1965)

VOL. 64, NO. 4 OCTOBER 1965

Psychological Bulletin

PHYSIOLOGICAL GRADIENTS AND BEHAVIOR[1]

ROBERT B. MALMO[2]

Allan Memorial Institute, McGill University

Physiological gradients accompanying mental activity have been found in skeletal-motor and autonomic recordings, commencing with the onset of the behavior sequence and terminating at its conclusion. Experimental evidence is presented indicating that these gradients do not signify increasing activation (or arousal) during the behavioral sequence (e.g., task or period of attentive listening). On the contrary, the EEG evidence clearly indicates that cortical activity remains relatively constant during the sequence when skeletal-motor and cardio-respiratory levels show progressive rise. While the gradients therefore appear not to represent increasing motivation, or the like, during the task, there is strong evidence indicating that the steepness of the gradients is a function of motivational level.

There is now extensive evidence from a wide variety of experimental situations for physiological gradients commencing with the onset of a continuous sequence (such as the performance of a tracking trial, listening to a story, or participating in an interview) and continuing as progressive rise to the termination of the sequence (typical gradients are presented in Figures 1 and 3). One of the purposes of this paper is to review this evidence. Another purpose is to consider what these gradients may mean with regard to central nervous system (CNS) and autonomic nervous system (ANS) mechanisms underlying organized behavioral sequences. Early studies related to this topic have been reviewed by Bartoshuk (1955b). Experiments on gradients in our laboratory commenced with observations on psychiatric patients.

PHYSIOLOGICAL GRADIENTS IN CLINICAL RESEARCH

In the first study (Malmo, Shagass, & Davis, 1951), electromyographic (EMG) gradients were clearly evident in the tracings, showing progressively rising EMG voltage (i.e., increasing muscle tension) from the onset to the termination of a mirror drawing test, usually with marked fall in tension at the completion of the drawing. These gradients were so very striking in the muscle potentials that (as will be obvious later in this paper) for some five years experiments designed to study physiological gradients in relation to performance (later with nonpatient subjects) concentrated on muscle potentials. To anticipate a later section of this paper, when a number of other physiological tracings were included in this kind of study (beginning with the one by Malmo & Davis in 1956), most showed gradients, so that in the discussion of EMG gradients that follows it should be kept in mind that these gradients are not peculiar to the skeletal motor system.

It may also be noted that even prior to 1956 Malmo and Shagass (1952) had found that psychoneurotic subjects showed a blood pressure gradient during a mirror drawing task (not shown by normal controls), and similar findings were also reported in this article for blood pressure change during a series of paced size discriminations.

There were also a number of EMG gradients that appeared in these clinical studies. Davis and Malmo (1951) obtained gradients from the *frontalis* muscle during interviews lasting an hour or more in a patient whose chief complaint was severe headaches which she referred to as bandlike sensations encircling the head just above the eyes and most intense over frontal and temporal regions. In one especially stressful interview the tension in the forehead reached a particularly high level, and the patient actually complained of headache. Supportive measures introduced by the therapist at this point were followed by a sharp drop in forehead tension.

---

[1] This work was supported in part by United States Public Health Service grants, a United States Army contract, and grants from the National and Medical Research Councils and the Defence Research Board of Canada.

[2] Thanks are due to A. K. Bartoshuk, P. F. MacNeilage, and P. M. Milner for critical reading of the manuscript.

In a case with severe tensional discomfort at the back of the neck, Case 2 in Shagass and Malmo (1954), neck EMGs showed a gradient, again with complaint of headache when the gradient had nearly reached its highest level. EMGs from the limbs (combined) also showed a gradient, but *frontalis* EMGs failed to do so. The agreement of locus of the somatic change with the patient's subjective report (e.g., *frontalis* EMG gradient in the patient with complaint of discomfort in the forehead and neck EMG gradient in the patient with complaint of headaches in the neck) is evidence for the "principle of symptom specificity" which has been stated in general terms by Malmo and Shagass (1949) and which is closely related to the principle of autonomic response stereotypy stated by Lacey and his co-workers (for a review, see Lacey & Lacey, 1958). Other examples of cases of this kind are cited in Malmo (1959, see Fig. 6, p. 382, and accompanying text for description of a symptom-specific EMG gradient that appeared during a tracking task carried out under stressful conditions).

GRADIENTS DURING CONTINUOUS LISTENING UNDER CONTROLLED EXPERIMENTAL CONDITIONS

In recording physiological reactions from psychiatric patients during interview there are a number of uncontrolled variables. First of all, with psychiatric patients as subjects it could have been that the gradients were simply another manifestation of psychoneurotic overresponsiveness (see Malmo, 1957, 1959; Martin, 1961); in the second place, the gradients could have been due in large part at least to some kind of cumulative effect of very stressful interview content.

That such conditions were not essential for the appearance of *frontalis* EMG gradients was demonstrated in an experiment by Wallerstein (1954). This investigator employed normal subjects who listened to an interesting detective story presented by a tape recorder, and during the period of listening (about 10 minutes), *frontalis* and chin tension showed progressive rise, falling at the conclusion of the story. In a second part of the same experiment the subjects listened to an excerpt from Kant's *Critique of Pure Reason*, and again during listening *frontalis* EMG gradients were obtained, though they were smaller in magnitude than those accompanying the detective story, and tension did not fall abruptly upon its termination, as was the case at the end of the detective story.

No significant increase in forearm muscle tension was found in listening to the detective story, but during the first hearing of the essay there was significant tensional rise in the forearm. Wallerstein cites similar findings in an earlier study by Smith, Malmo, and Shagass (1954) in which tension in the arms rose significantly during listening to a relatively uninteresting article which had purposely been made difficult to hear in places by decreasing the volume control of the recording apparatus, at intervals, to a point where subsequent reproduction was almost inaudible—although it could be heard if the subject made an effort. In both articles the significantly increased forearm muscle potentials were interpreted as reflecting effortful listening. "Straining" to hear, in which use is made of the skeletal muscles, is a commonly observed phenomenon, but to the best of this author's knowledge no clear formulation has ever been made concerning what purpose the muscles can serve in such circumstances. In dealing later with interpretations of gradients in general, some light may be thrown on this question.

In his interpretation of EMG gradients from facial (chin and *frontalis*) muscles, Wallerstein suggested that these gradients may have been "associated with increasing comprehension or organization of incoming verbal material—organization which may take place during attentive listening [Wallerstein, 1954, p. 238]." It is important to note that in his interpretation of the chin and *frontalis* muscle gradients, Wallerstein follows Hebb's earlier approach to motivation (1949, p. 228 ff.) prior to the formulation of his "arousal" concept (Hebb, 1955).

A follow-up experiment by Bartoshuk (1956) again employed the same detective story with normal subjects listening. But in addition to EMGs, frontal and parietal EEGs were also recorded. Combined results from the Wallerstein and Bartoshuk experiments strengthened the demonstration of significant forehead EMG gradients during listening. Positive correlations between gradient steepness and reported interest were interpreted by Bartoshuk as supporting his hypothesis (different from Wallerstein's, 1954) that steepness of forehead EMG gradients "were indicative of motivation to listen." In the first hearing of the story (but not in subsequent repetitions of it) the steepness of forehead EMG gradients was inversely related to EEG amplitude. This relation between EMG gradients and EEG on the first hearing was interpreted in terms of Hebb's (1955) "arousal function," the steeper the gradient the higher the arousal. Bartoshuk (1956, p. 161 f.) explained that it is the gradient as a whole which is related to arousal level, and he did not conclude that the rising tension during listening was paralleled by growing interest in the story, pointing out that to test this it would be necessary to divide the story into several sections, each of which could be rated for interest.

Furthermore, Bartoshuk demonstrated with quantified EEG data (in which the alpha component was most prominent) that the mean EEG voltages themselves failed to show any gradients as would have been the case with progressively greater cortical activation during the period of listening.

Lack of correlation between EMGs and EEGs on repeated presentations of the story was interpreted in accordance with the terminology of Hebb (1955) as indicating that arousal function was less a factor in repeated hearing and that the gradients on later hearings might be understood entirely in terms of cue function. The nature of this cue function effect was not specified. Bartoshuk suggested that further analysis of relations between cue and arousal functions might be more profitably done with tasks in which performance level could be used as a measure of cue function and in which the subject's motivational level ("arousal function") could be more easily subjected to experimental manipulation.

Physiological Gradients in Perceptual Motor Tasks

Again, for summary of the early work related to this topic see Bartoshuk (1955b). Following the work in our laboratory with mirror drawing in psychiatric patients there was a series of studies with normal subjects beginning with Smith's (1953) experiment. Using mirror drawing, Smith obtained EMG gradients confirming the findings of Malmo, Shagass, and Davis, (1951). Smith's interpretation of these gradients in principle seems very similar to Wallerstein's (1954) interpretation of gradients during attentive listening. Smith's hypothesis was stated as follows:

> the observed increase in muscle tension during drawing reflects the developing organization of a central neural process, which controls and "anticipates" the overt activity, and is yet relatively independent of immediate sensory stimulation [1953, pp. 34–35].

Bartoshuk found significant relations between steepness of EMG gradients with amount of increase in residual tension following task interruption (1955a) and with speed and accuracy of mirror drawing (1955b), which he interpreted as supporting a motivational hypothesis of EMG gradients. Surwillo (1956), employing a visual tracking task, showed that by varying incentive he could alter the steepness of EMG gradients. Moreover, by using a task that was invariant in trial duration he introduced a control that was absent in Bartoshuk's experiments where trial duration varied according to drawing speed.

Surwillo also brought some support to Bartoshuk's notions concerning EMG gradient steepness being related to level of performance. Data from subjects performing two tracking trials under conditions of very high incentive were fractionated in the case of each subject according to whether the tracking score was the better or the worse for these two trials. EMG gradients for the left forearm muscles (i.e., from the arm not engaged in tracking) turned out as predicted with the steeper mean gradient corresponding to the better performance. Since this positive finding appeared in only one of the five muscles studied, Surwillo did not feel that the Bartoshuk hypothesis was strongly substantiated. However, the difference in incentive between the two trials which he selected for study was probably not very great (this kind of comparison will be dealt with later when some recent unpublished data are presented), and the particular muscles showing the most pronounced gradients appear to vary from one kind of task or experimental situation to another. For example, forehead EMGs that showed the best gradients during listening have been found inferior to other muscles in showing gradients in mirror drawing (see Bartoshuk, 1955b, p. 219; Malmo and Davis, 1956, Table I).

A very important observation made by Surwillo (1956) and later confirmed by Stennett (1957) was that grip pressure on the knob used in tracking actually fell during tracking. Therefore, the EMG gradients cannot be accounted for merely in terms of progressively increasing muscular exertion (in the context of force applied to the knob) during the task. This finding appears to have been overlooked by Eason (1959) and by Eason and Branks (1963) in their interpretation of EMG gradients. Eason (1959) failed to obtain EMG gradients under conditions in which force applied to a dynamometer declined during the trial. However, his failure to obtain gradients was very likely due to the simplicity (the low "mental" demands) of the task (squeezing a hand dynamometer). Here, Bélanger's (1957) findings appear relevant: Bélanger's subjects in a task requiring a series of rapid visual discriminations were required to press a button with their index fingers each time a judgment was made. Under these conditions EMG gradients were obtained. But when all they had to do was to press the button at the same intervals *without* being required to make any discriminations, no EMG gradients appeared. Eason's (1959) findings appear similar to these control results of Bélanger; and Bartoshuk's interpretation concerning EMG gradients and motivation therefore appears to be better supported by the facts than the opposing interpretation of Eason and Branks (1963, p. 541), which holds that the progressive increase in EMG actually signifies increasing level of activation.

That gradients were not confined to EMGs but occurred also in other physiological functions seemed a reasonable inference from the earlier work with psychiatric patients. This was confirmed with normal subjects in a mirror-drawing experiment by Malmo and Davis (1956). Gradients were found in heart rate (HR), blood pressure, and respiration in addition to those in muscle tension.

Finally, some unpublished tracking data, in addition to providing some new information, confirm the Malmo and Davis (1956) findings of HR and respiration gradients and Bartoshuk's (1956) failure to obtain EEG gradients. A full description of the experiment from which these data were taken is presented elsewhere (Malmo, in press). The tracking apparatus has been described by Davis and Mundl (1964). Figure 1 presents EMG gradients for the left *pronator teres* muscle, showing a gradient steeper for the best trial in four than for the worst trial. These four trials came from the so-called unified set condition in the experiment by Malmo (in press). The difference between mean left *pronator teres* EMG levels for the low and high distance-off-target (error) data was significant ($p < .025$, one-tailed).

Figure 2 presents the corresponding performance data. Note that in contrast to the regular rising gradients in Figure 1, the error curves show fluctuations and that the low error curve, at least, fails to show any overall gradient. The three peaks in error scores occur close to the 24-second point in the tracking cycle which would correspond to the place where the subject, having just reached the stop, was forced to reverse direction in rotating the tracking knob (from counterclockwise to clockwise). It is clear from these data that the regularly rising curves in Figure 1 cannot be accounted for in terms of increasing or decreasing mean error (i.e.,

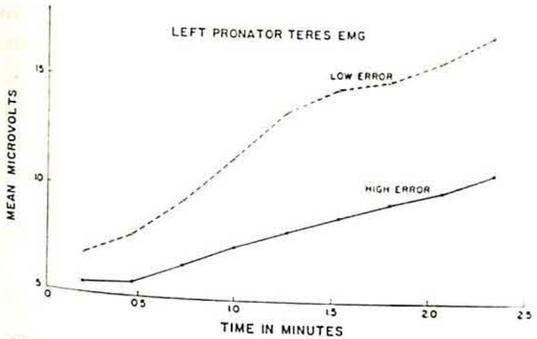


FIG. 1. Comparison of left *pronator teres* EMG gradients for high- versus low-error trials in manual tracking. (Numbers on the baseline represent the 4-second integration intervals that were selected for measurement. Each point is based on the mean of measurements from 35 subjects. Data from 7 subjects were unavailable for this analysis.)

in terms of any progressive change in task difficulty within trials).

Figure 3 presents gradients for respiration and HR. The HR curves in particular are very regular and fail to show any trace of irregularity corresponding to the fluctuations in the error curves in Figure 1. The difference between mean HR for low- and high-error conditions was marginally significant ($p < .05$, one-tailed). The corresponding difference for respiration was significant ($p < .025$, one-tailed).

Figure 4 presents curves for EEG in the 17–30 cycles per second (beta range) plotted from the same integration intervals as those used in Figures 1 and 2; and the absence of a regularly rising gradient is quite evident.

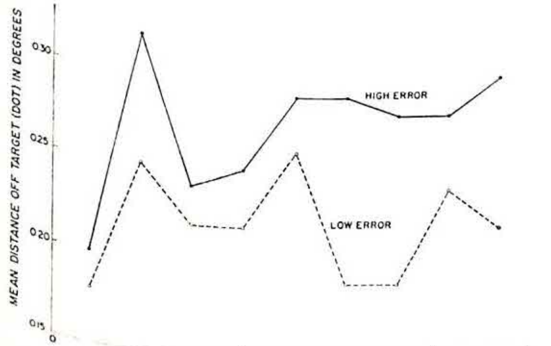


FIG. 2. Mean error curves are plotted from same 4-second integration samples as those used for plotting the curves in Figure 1. (These data are based on mean distance-off-target (DOT) scores from 40 subjects. Data from 2 subjects were unavailable for this analysis.)

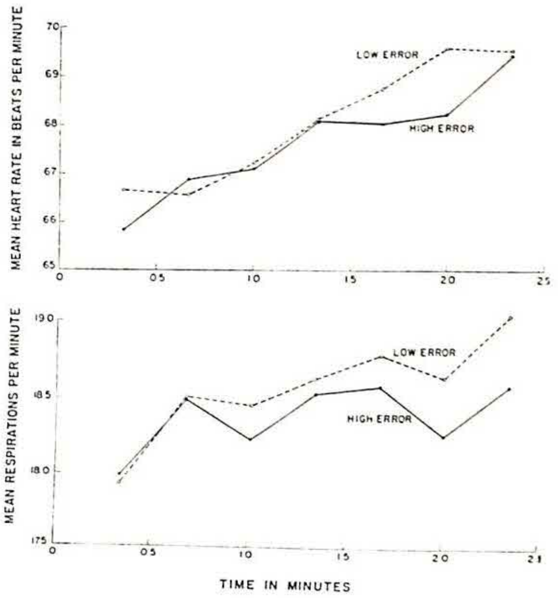


FIG. 3. Gradient curves for HR ($N = 40$) and respiration ($N = 35$). (Points on the base line are based on means for successive sevenths of the tracking trial. Data for 2 subjects were unavailable for HR, and data for 7 subjects were unavailable for respiration analyses.)

As previously mentioned, Bartoshuk (1956) showed by means of quantified data that EEG gradients failed to appear in a listening situation in which forehead muscle gradients clearly appeared. Bartoshuk's filters were so constructed that the high voltage alpha frequencies dominated the picture. The data in Figure 4 provide further negative data, this time showing absence of gradients in EEG data from the beta range. Furthermore, the results of quantitative analysis of EEG alpha from the present study (in which an 8–12-cycles-per-second filter was employed) confirmed Bartoshuk's earlier finding with regard to absence of EEG alpha gradients. In the high error condition, 21 of 40 subjects (i.e., only 52.5%) showed an increase in voltage from the first to last point on the curve. Under the low error condition there were 18 of 38 (only 47.37%) of the subjects showing such an increase. These percentages are clearly not significant from chance (50%).

There appears, therefore, to be no instance of EEG voltage in any of the frequency bands studied showing progressively increas-

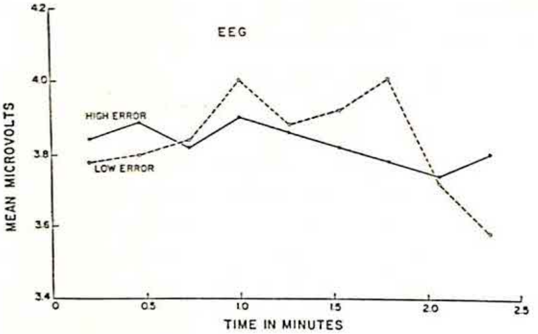


FIG. 4. Curves for mean 17–30 cycles per second. (EEG values are plotted in the same way as those for the data in Figures 1 and 2. Each point is based on data from 35 subjects. Data for 7 subjects were unavailable for this analysis.)

ing activation. This should not be particularly surprising since there is no evidence on other grounds for interpreting the gradients as signifying gradually rising activation level. That steepness of gradient and EEG indices of activation are significantly correlated under certain circumstances was demonstrated by Bartoshuk (1956), who, as previously mentioned, found very significant inverse relations between mean EEG voltages and forehead EMGs during the first hearing of a story. This of course means that during the first hearing, subjects with steep EMG gradients showed the typical, low-voltage, "activated" EEG, whereas subjects in whom such steep gradients failed to appear showed the typical "nonactivated" EEG pattern (see Bartoshuk, 1956, Figure 3). Absence of correlations between EMG and EEG on later hearings of the story was interpreted by Bartoshuk as indicating that gradient steepness was multiply determined, that activation level was an important factor in determining gradient steepness on the first hearing of the story but not a significant factor when the same story was repeated immediately afterwards. In suggesting what one of these other factors might be, Bartoshuk singled out the cue function as postulated by Hebb (1955).

EEG data in the study herein reported (see Figure 4) did not show a significant difference between high- and low-error conditions. It should be pointed out, however, that the mean difference between these two conditions was relatively slight. For low error the mean off-target score was .21 degree compared with .28 degree mean score for high error, a $33\frac{1}{3}\%$ increase. Compared with this, the difference between single and double tracking was much greater: .25 degree mean score for single and .42 degree for double tracking, an increase of 68% (see Malmo, in press). Here it should be stated that because of the relatively slight differentiation provided by the high- versus low-error comparison, no very large and significant differences should have been expected between conditions for the peripheral physiological functions (EMG, HR, and respiration) which showed gradients.

The chief importance of this study was in showing that with no significant intratrial trend in performance, and in the presence of rising EMG gradients, there were no EEG gradients. Interesting trends may appear in EEG over time as MacNeilage[3] has shown, but these trends in EEG observed by MacNeilage were not found associated with EMG gradients so that there is really no conflict between his findings and mine.

If it is steepness of the gradient that, under certain circumstances, gauges level of activation within a performance trial (or during listening and the like) and it is *not* the case that activation increases *pari passu* with the EMG gradients, then there is no reason really to expect EEG to show a gradient. As a matter of fact it appears that gradients, when observed at all, seem to appear in measures of "background" effector action, and EEGs, of course, do not fit into this category.

Of 17 studies, where they were looked for, EMG gradients were found in 16 (Bartoshuk, 1955a, 1955b, 1956; Bélanger, 1957; Davis & Malmo, 1951; Elliott, 1964; Malmo, in press, 1959; Malmo & Davis, 1956; Malmo & Shagass, 1952; Malmo, Shagass, & Davis, 1951; Malmo & Surwillo, 1960; Smith, 1953; Stennett, 1957; Surwillo, 1956; Wallerstein, 1954). Failures to observe EMG gradients in adults have been relatively rare. MacNeilage (see Footnote 3) recorded forehead and right *pronator teres* EMGs in his study, and he observed no gradients from either muscle during a paced auditory serial-addition task. It would be of interest to repeat this experiment, recording from 6 to 8 muscles, in order to determine whether some muscles other than the ones recorded from in MacNeilage's experiment might show gradients. The subjects used their right hands to write down their serial additions, and from other studies (see, for example, Stennett, 1957, Figures 2 and 3) it seems that EMGs from the left ("passive") arm would be promising ones to examine for gradients.

---

[3] P. F. MacNeilage. Personal communication.

Failure to obtain EMG gradients in young children under experimental conditions where typical gradients were observed in adults was recently reported by Elliott (1964, p. 24). From these and other data Elliott concluded "that the muscle tension response in 6-year-old children does not have the same motivational significance that it has in adults."

The effect of repeated sessions on gradients is in need of further study. From the results of an experiment on sleep deprivation by Malmo and Surwillo, it appears that in the absence of exceedingly strong incentive, gradients from certain muscles may decline in steepness during tracking trials (see Malmo & Surwillo, 1960, Figure 6, p. 13). That gradient steepness does not simply decline with long continued practice on a motor task was shown by Bartoshuk (1955b, p. 219), for mirror drawing, and by Malmo and Surwillo (1960, p. 19) in their findings of EMG response specificity during sleep deprivation.

PALMAR CONDUCTANCE (PC)

In contrast to gradients for other physiological measures, gradients for PC are falling ones. The typical finding is illustrated in Figure 1 of Stennett's (1957) article. The strongest PC reactions occur between trials in tasks such as tracking. During the trial there is typically a slow return to base line that produces a falling curve. Generally speaking, PC appears very sensitive to transitions, that is, to shifts in the kind of activity being performed or to sudden changes in environmental stimulation. For recent corroborations of this point see Elliott (1964), and Harter, Eason, and White (1964) who cite earlier studies also confirming this general conclusion.

POSSIBLE NEURAL MECHANISMS UNDERLYING PHYSIOLOGICAL GRADIENTS

At the present stage of knowledge only very tentative suggestions can be made in attempting to account for the observed phenomena. As this review has indicated, the observations themselves have on the whole been consistent and reproducible (surprisingly so in view of the rather different situations in which they have been made). To summarize the main findings, progressively rising gradients have been observed in recording skeletal muscle tension (often from muscles not engaged in activity) and in recording cardiovascular and respiratory changes. These gradients have been found to accompany periods of sustained attention, commencing with the onset of the period and rising monotonically to the end, usually with a pronounced drop at the end. The time courses of the gradients have depended on the duration of the behavioral sequences, from 20–30 seconds in the case of some mirror drawing trials, for example, up to 10 minutes in the case of hearing a story presented by a tape recorder (and, as previously stated, even longer gradients, up to an hour or more, have been obtained from interviews with psychiatric patients).

Again, the evidence indicates that the rising gradients do not signify *increasing* motivation, degree of interest, or the like, *during* the course of the behavioral sequence. Instead, the indications are that this increasing "tonic" background is somehow part of the whole physiological activity seemingly required for sustaining a relatively even level of attention from the beginning to the end of the task (or other sequence of mental activity).

At the present stage of neuropsychological knowledge, going beyond the statement contained in the last paragraph is frankly speculative. However, in order to advance on the neuropsychological problem, a certain amount of speculation is required (to bring further experimental work into focus).

The theoretical discussion by Miller, Galanter, and Pribram (1960) provides a good point of departure. They said: "*A Plan is any hierarchical process in the organism that can control the order in which a sequence of operations is to be performed* [p. 16]." They went on to say: "The execution of a Plan need not result in overt action—especially in man, it seems to be true that there are Plans for collecting or transforming information, as well as Plans for guiding actions [p. 17]." There seems little doubt that the unit of behavior which they described is identical with the unit which in our studies is demarcated by onset and termination of physiological gradients.

An interesting parallel between the Miller, Galanter, and Pribram analysis and ours comes out in their discussion (p. 66) of the consequences of task interruption. This discussion serves also to bring into the same framework the important work of Lewin (1935) and his associates. Early in the series of experiments on the physiological gradient phenomenon Smith (1953) found in a mirror drawing task that muscle tension dropped significantly more after completion than after interruption of the task. This finding fits nicely with Lewin's hypothesis that task interruption serves to continue or prolong a "tension system." Furthermore, there is another important area of agreement between the viewpoints of Lewin, Miller, Galanter, and Pribram and ourselves. In order for a tension system or "physiological gradient" to manifest itself, the task cannot be too simple and repetitious. For physiological gradients this was clearly shown in the aforementioned experiment by Bélanger (1957). On this main conclusion there is agreement between the three points of view, though it should be pointed out that there is not complete agreement between prediction from Lewinian theory and from the theory of Miller, Galanter, and Pribram (1960, p. 66).

Again, on a frankly speculative basis (but drawing on relevant facts where they exist), we may raise questions concerning what neurophysiological mechanisms could be reflected in the physiological gradients that we have been observing. As a starting point we may note the conclusion of Miller, Galanter, and Pribram (1960) that:

> Neurobehavioral studies conducted on animals support . . . [the] notion that the limbic systems of the internal core of the forebrain play an essential role in the execution of Plans [p. 204].

They went on to make further, more detailed suggestions (admittedly speculative):

> The hippocampus could perform . . . [a] "gating" function via the reticular formation in the internal core of the brain stem, which, in turn, is known to influence the receptors, the afferent pathways into the central nervous system, and the activities of the entire external portion of the forebrain. Thus the hippocampus may be intimately involved in the business of keeping the brain at work on the successive steps in the Plan . . . [pp. 206–207].

To bring the discussion back to the data on physiological gradients, the EEG evidence indicates that cortical activity is relatively constant while cardio-respiratory and EMG activity is progressively rising during the task or other behavioral sequence.

Probably the most likely notion is that adaptation or fatigue is responsible for this differential reaction, at least in part.[4] It is not unreasonable to assume that cortical neurons and circuits are more susceptible to adaptation than the more peripheral parts of the motor system (for a related inference see Hebb, 1949, p. 226). Then, if there is a homeostatic feedback mechanism that keeps activation level relatively constant during a sequence where sustained attention is required, this will involve gradually rising activity of part of the arousal system to overcome the reduced sensitivity of the "higher" circuits. The affluxive part of this mechanism would presumably be the ascending reticular activating system (ARAS). From the physiological (gradient) evidence, it would seem that the descending branch of the reticular system is also brought into play. Bearing in mind the probably greater sensitivity of cortical circuits, if the same arousal signal is fed via ARAS and also downward via the descending reticular facilitatory mechanism in the brain stem, this signal (while maintaining a relatively constant condition in the cortex) will produce a gradually increasing peripheral tonus because of the lesser decay in sensitivity to bombardment in the more peripheral parts of the motor systems.

---

[4] Thanks are due P. M. Milner for suggesting this idea to me.

Because EMG gradients have been so frequently found in passive muscles (see Figure 1, for example), the question arises whether an alternative fatigue concept might be proposed in which gradients would be produced only in unused muscles. However, quite marked EMG gradients have been recorded from the active arm in motor tasks (see, for example, Bartoshuk, 1955a, 1955b; Surwillo, 1956), ruling out any general application of such an explanation.

Regarding details of the more peripheral mechanisms, Bartoshuk (1955b, 1964) has stressed the probable importance of supraspinal stimulation of muscle-spindle feedback circuits in accounting for EMG gradients. Supporting this idea is the finding of extensive distribution of sensitive foci in the reticular formation capable of influencing muscle spindles in the cat's hindlimb (Eldred & Fujimori, 1958). Furthermore, with regard to the notion advanced earlier concerning the peripheral part of the system being less fatigable, it has long been known that the muscle spindles are relatively nonadapting (Fulton, 1949, p. 16).

While up to this point the discussion has dealt specifically with effects on circuits for control of skeletal-motor activity, the main arguments introduced appear applicable to the autonomic effector mechanisms (for a discussion related to this problem see Malmo, 1963, p. 22 ff.).

REFERENCES

Bartoshuk, A. K. Electromyographic gradients in goal-directed activity. *Canadian Journal of Psychology*, 1955, **9**, 21–28. (a)

Bartoshuk, A. K. Electromyographic gradients as indicants of motivation. *Canadian Journal of Psychology*, 1955, **9**, 215–230. (b)

Bartoshuk, A. K. EMG gradients and EEG amplitude during motivated listening. *Canadian Journal of Psychology*, 1956, **10**, 156–164.

Bartoshuk, A. K. Excitability of stretch receptors and alpha efferents during simple tone discrimination. *Perceptual and Motor Skills*, 1964, **18**, 523–524.

Bélanger, D. "Gradients" musculaires et processes mentaux supérieurs. *Canadian Journal of Psychology*, 1957, **11**, 113–122.

Davis, F. H., & Malmo, R. B. Electromyographic recording during interview. *American Journal of Psychiatry*, 1951, **107**, 908–916.

Davis, J. F., & Mundl, W. J. A dual modality tracking system. *Psychophysiology*, 1964, **1**, 183–191.

Eason, R. G. The surface electromyogram (EMG) gauges subjective effort. *Perceptual and Motor Skills*, 1959, **9**, 359–361.

Eason, R. G., & Branks, J. Effect of level of activation on the quality and efficiency of performance of verbal and motor tasks. *Perceptual and Motor Skills*, 1963, **16**, 525–543.

Eldred, E., & Fujimori, B. Relations of the reticular formation to muscle spindle activation. In H. H. Jasper, L. D. Proctor, R. S. Knighton, W. C. Noshay, R. T. Costello (Eds.), *Reticular formation of the brain.* Boston: Little, Brown, 1958. Pp. 275–283.

Elliott, R. Physiological activity and performance: A comparison of kindergarten children with young adults. *Psychological Monographs*, 1964, **78**(Whole No. 587), 1–33.

Fulton, J. F. *Physiology of the nervous system.* (3rd ed.) New York: Oxford, 1949.

Harter, M. R., Eason, R. G., & White, C. T. Effects of intermittent visual input disruption, flicker-rate, and work time on tracking performance and activation level. *Perceptual and Motor Skills*, 1964, **19**, 831–848.

Hebb, D. O. *Organization of behavior.* New York: Wiley, 1949.

Hebb, D. O. Drives and the C.N.S. (Conceptual Nervous System). *Psychological Review*, 1955, **62**, 243–254.

Lacey, J. I., & Lacey, Beatrice C. Verification and extension of the principle of autonomic response-stereotypy. *American Journal of Psychology*, 1958, **71**, 50–73.

Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

MacNeilage, P. F. Changes in EEG and other physiological measures during serial mental performance. Unpublished doctoral dissertation, McGill University, 1961.

Malmo, R. B. Cognitive factors in impairment: A neuropsychological study of divided set. *Journal of Experimental Psychology*, in press.

Malmo, R. B. Anxiety and behavioral arousal. *Psychological Review*, 1957, **64**, 276–287.

Malmo, R. B. Activation: A neurophysiological dimension. *Psychological Review*, 1959, **66**, 367–386.

Malmo, R. B. On central and autonomic nervous system mechanisms in conditioning, learning, and performance. *Canadian Journal of Psychology*, 1963, **17**, 1–36.

Malmo, R. B. Finger-sweat prints in the differentiation of low and high incentive. *Psychophysiology*, 1965, in press.

Malmo, R. B., & Davis, J. F. Physiological gradients as indicants of "arousal" in mirror tracing. *Canadian Journal of Psychology*, 1956, **10**, 231–238.

MALMO, R. B., & SHAGASS, C. Physiologic study of symptom mechanisms in psychiatric patients under stress. *Psychosomatic Medicine,* 1949, **11**, 25–29.

MALMO, R. B., & SHAGASS, C. Studies of blood pressure in psychiatric patients under stress. *Psychosomatic Medicine,* 1952, **14**, 82–93.

MALMO, R. B., SHAGASS, C., & DAVIS, J. F. Electromyographic studies of muscular tension in psychiatric patients under stress. *Journal of Clinical and Experimental Psychopathology,* 1951, 12, 45–66.

MALMO, R. B., & SURWILLO, W. W. Sleep deprivation: Changes in performance and physiological indicants of activation. *Psychological Monographs,* 1960, **74**(Whole No. 502), 1–24.

MARTIN, IRENE. Somatic reactivity. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1961. Pp. 417–456.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

SHAGASS, C., & MALMO, R. B. Psychodynamic themes and localized muscular tension during psychotherapy. *Psychosomatic Medicine,* 1954, **16**, 295–313.

SMITH, A. A. An electromyographic study of tension in interrupted and completed tasks. *Journal of Experimental Psychology,* 1953, **46**, 32–46.

SMITH, A. A., MALMO, R. B., & SHAGASS, C. An electromyographic study of listening and talking. *Canadian Journal of Psychology,* 1954, **8**, 219–227.

STENNETT, R. G. The relationship of performance level to level of arousal. *Journal of Experimental Psychology,* 1957, **54**, 54–61.

SURWILLO, W. W. Psychological factors in muscle-action potentials: EMG gradients. *Journal of Experimental Psychology,* 1956, **52**, 263–272.

WALLERSTEIN, H. An electromyographic study of attentive listening. *Canadian Journal of Psychology,* 1954, **8**, 228–238.

(Received April 7, 1965)

Psychological Bulletin
1965, Vol. 64, No. 4, 235–245

EFFECTS OF CHLORPROMAZINE ON LEARNING

LAWRENCE C. HARTLAGE[1]

Department of Mental Health, Louisville, Kentucky

Learning is an essential part of behavior therapies and is an important requirement of more traditional psychotherapy. The widespread use of tranquilizing drugs in psychiatric settings prompts a review of their effects on learning. Chlorpromazine is perhaps the most widely investigated tranquilizing drug and is used as a referent in the assessment of its effect on learning. Results of studies involving a number of animals, normal subjects, and psychiatric patients tend to show significant declines in learning on a wide range of tasks, with a linear decline in learning with increased dosage levels. A number of methodological difficulties which limit generalizations are discussed.

The widespread use of tranquilizing drugs in psychiatric settings (Baron & Fisher, 1962) makes relevant an investigation of the effect of these agents on the learning of the individuals being treated. In light of the impact of the conditioning therapies on the field of psychotherapy (Wolpe, Salter, & Reyna, 1964), and the inevitability of the need for learning of some sort to take place during more traditional forms of psychotherapy (Munroe, 1955), the conditionability of the individual treated by any psychotropic agent becomes an important factor in the outcome of therapy. The first published report in English on the effectiveness of the psychotherapeutic effectiveness of a tranquilizing drug involved chlorpromazine (Delay, Deniker, & Harl, 1952), and by 1956 nearly 1,000 articles on this drug had been published (Bennett, 1957). Since that time, nearly 1,200 articles dealing with chlorpromazine have appeared in the *Psychopharmacology Abstracts*, making this drug perhaps the most investigated single psychotropic agent (Freyhan & Merkel, 1961). As Ostow (1962, p. 300) points out, it has become a standard against which all others are compared.

METHODOLOGICAL CONSIDERATIONS

The importance of methodological issues in the field of psychopharmacology has been emphasized as essential to the future of the discipline (Cole & Gerard, 1959), and methodological issues have figured prominently in recent publications in this field (Dews & Morse, 1961; Hunt, 1960; Uhr & Miller, 1960). Unfortunately, this has not always been the case: when Bennett (1957) reviewed nearly 1,000 published articles dealing with psychopharmacology, he concluded that only 10 had employed proper controls. In light of inconclusive evidence concerning locus of action, global neurochemical effects, and specificity of function, perhaps the most fruitful approach to the problem at this time may lie in the investigation of specific behavioral indices of learning related to administration of the drug.

As Russell (1964) has pointed out, it is quite possible to study empirical relations between input, that is, chemical stimulation via drug action, and output, that is, changes in behavior, disregarding the intervening events. This stimulus-response type of operational relationship permits generalizations about the effect of a given drug on a specified behavior without the necessity of postulating any functional interactions of the biological systems involved. Within such an operational framework, it is possible to regard a drug as the independent variable and some measure of learning as the dependent variable.

Psychopharmacologists have typically recognized that the temporal order or sequential effects of drugs might influence their results, and so have often alternated or balanced the orders of presentations with a "crossover" design (Barmack, 1940). This design is a special case of the "Latin square" design and

---

[1] Now at Larue Carter Hospital, Indianapolis, Indiana.

My thanks to L. B. Miller and R. P. Smith for a preliminary reading of the manuscript.

requires that all interactions be zero, a requirement which, as McNemar (1961) points out, seriously restricts its use in psychopharmacology research. In the Latin square, therefore, all interactions are omitted from the analysis, and the effects of these interactions become indistinguishable from the variations of the response being studied and so become "confounded" with experimental error (Nash, 1960, p. 133). The requirement that all interactions be zero is probably seldom met, since, as Kornetsky, Humphries, and Evarts (1957) demonstrated, there is a typical Drug X-subject interaction due to personality differences in reactivity. Further, drug-induced changes in learning capacity may readily be confounded with drug effects within this paradigm. For example, the drug may impair the capacity of subjects who are first tested in the drugged state to understand the task set before them (Jellinek & McFarland, 1940).

Another major methodological problem is that of a suggestion factor. As Rosenthal and Frank (1956) point out, this is often sufficient to affect the outcome, and their findings have repeatedly been supported (Beecher, Keats, Mosteller, & Lasagna, 1953; Jellinek, 1946; Lasagna, Mosteller, von Felsinger, & Beecher, 1954). Although the "double blind" model provides a control for this, there are typically enough cues from side effects to transcend such controls (Baker & Thorpe, 1957).

The problem of response sets is one perhaps related to that of suggestibility. As Fleischman (1953) and Fleischman and Hempel (1954, 1955) have pointed out, these factors underlying test performance may shift during the course of testing, making any findings equivocal and inconclusive. None of the measures of learning in the field of psychopharmacology appear to have exerted any controls over this problem, so the extent to which findings may be confounded by response sets is moot.

RESULTS

Although superficially it would appear that there is a great deal of discrepancy among reported results of the effects of chlorpromazine on learning, a close examination of these reports typically indicates that the apparent discrepancies are due to lack of specification or control of one of the parameters involved. For example, some investigators (Kovitz, Carter, & Addison, 1955; Shaten, Rockmore, & Funk, 1956) studying the effects of chlorpromazine in schizophrenic patients have generally concluded that chlorpromazine either improves, or does not significantly impair performance. Kornetsky's (Kornetsky, Humphries, & Evarts, 1957) contrary results with similar tests of performance in normals were, after further investigation, found to represent differences in dosage levels, chronic versus acute administration of the drug, and variations between various types of subtests as well as differences between normals and schizophrenics (Kornetsky & Humphries, 1958; Kornetsky, Pettit, Wynne, & Evarts, 1959).

Animal Conditioning. Conditioned responses have been found to be impaired in a wide variety of infrahuman species, ranging from pigeons to monkeys, under various dosage levels of chlorpromazine. Cook, Kelleher, and Fellows (1962), using dosages of 5–30 milograms per kilogram of body weight (mg/kg), found depressant effects on key pressing in pigeons. Ferster, Appel, and Hiss (1962), with dosages of 1–10 mg/kg in pigeons, found a decrease in response rate at higher dosages but an increase in lower and middle dosages. Rahmann (1962) demonstrated that with hamsters dosages of .5, 1.0, and 2.0 mg/kg impaired conditionability significantly more than a placebo administered to a control group; and even after learning to criterion, there was much less resistance to extinction in spite of the greater number of reinforced trials. Isolated studies with cats (Pautler & Clark, 1961), dogs, and rabbits (Sadowski & Longo, 1962) produced similar decrements in conditioned responses. Dosages ranged from 1.9 mg/kg (Paulter & Clark) to 225 mg. Working with mice, Jacob and Blozovski (1961) identified another relevant variable. Using dosages of 1.6, 2.5, 2.8, 3.0, and 3.5 mg/kg, they found significant impairment of time required for conditioning, and with dose levels of 10 mg/kg, quantity of conditioning was also decreased. They further found that conditioned responses were deleteriously affected 24 hours after dosing, even when they had not been affected at the time of administration of the drug. Rosenberg and Savaris (1963) found decrements in conditioned responses one hour after dosing at 12.5 mg/kg, and Clark (1962), using levels ranging from 16 to 45 mg/kg, found severe impairment in conditioned motor responses. This may be due to the rather high dosage levels used, however, since White and Westerbeke (1961) found a change in EEG alert patterns to deactivation patterns with high concentrations of the drug, and similar EEG results were reported by Everett (1961) and Paillard, Bert, Zwingelstein, and Giudicelli (1961). Further, Posluns (1962) found a motor deficit at higher dosage levels which may have accounted for the lack of responding, and Wolf, Swinyard, and Clark (1962) reported similar findings.

As might be expected, perhaps the largest body of reported research on subhuman species has been done with rats. Gatti and Frank (1961) found that a 2-mg/kg dose increased the latency of a conditioned reflex 30 minutes after administration, abolished it 2 hours after administration, and returned it to near normal after approximately 6 hours. Irwin and Armstrong (1961) found that conditioned locomotor activity on a treadmill was significantly depressed by chlorpromazine and showed an increasing depression over continued administration. Finocchio (1963) reported similar results with a lever-pressing conditioned response. The dimension of duration of onset was demonstrated by Izquierido (1962), using 2.0–2.5 mg/kg injected intraperitoneally (ip). He found an inhibition of conditioned responding within 10–20 minutes. Boren (1961) found a consistent depression in conditioned responses with 1–3 mg/kg injected intraperitoneally 10 minutes before the initiation of the session. Heimstra (1962), using oral administration, found no suppression in conditioned response after 2 hours, but Gonzales and Shepp (1962) found 4 mg/kg (ip) 2 hours and 4 hours before testing had completely supressed the conditioned responses. The rate of onset appears to be substantially faster for ip administration. Murphree, Monroe, and Seager (1962) found a significantly lower survival rate among infant rats whose mothers had been given 5 mg/kg of chlorpromazine regularly during pregnancy. It is difficult to say whether this was due to a physiological factor or a loss of adaptive behavior, although human infants whose mothers had been on comparable dosages during pregnancy showed no conclusive evidence of physiological changes (Kris, 1962; Werboff & Kesner, 1963). Bindra and Mendelson (1963) found differential effects of chlorpromazine on higher and lower levels of conditioning. He reported a decrease in response rate at higher levels of training, but not at the highest level. Aaronson (1963) has criticized the commonly used pass-fail criteria, on the ground that it would be difficult to demonstrate changes at the extremes of response measurements; Bindra's results may be somewhat affected at the highest level by this criterion artifact. Plotnikoff (1961) found that inbreeding increased resistance to the effects of chlorpromazine on conditioning measures, and Aceto, Lynch, and Thomas (1961) found that fasted rats produced greater reactivity to the drug, making studies which do not control for these variables open to question concerning the effect these factors might have exerted on outcomes.

Schieckel (1962) found a curvilinear relationship between dosage levels and conditioned delay behavior in monkeys. Using a range of from .31 to 1.0 mg/kg, he found large doses produced a significant decrease in the limit of delay, whereas small and medium doses resulted in increases in limit of delay. Weitzman and Ross (1962) found that .3 mg/kg produced marked fluctuations in responses, and Clark, Jackson, and Brady (1962) reported a general decline in conditioned behavior after administration of chlorpromazine.

A class of conditioned response of special interest to psychopharmacologists is the conditioned emotional response. The conditioned emotional response has certain face validity as an animal analog of human psychiatric disorder (Hunt, 1960). Through learning, a neutral stimulus acquires the power to evoke an intense emotional disturbance, one greater than that which the reinforcing shocks could evoke initially (Estes & Skinner, 1941). The ease of establishing the conditioned emotional response (CER) and of controlling its strength experimentally, together with an ostensible relevance to human emotional disorders, has favored its use as an indicator in a number of studies of conditioning and of the effects of tranquilizing drugs on conditioning.

Hunt (1956) and Jernberg (1957) have found that heavy doses of chlorpromazine have marked effect on both expression and extinction of the CER, but interfere only slightly with acquisition. Halasz (1959) however, found that moderate doses facilitate the extinction of avoidance and conditioned fear in the cat. Appel (1963) found that the conditioned emotional response in rats was inhibited with a dosage of 1.5 mg/kg, and Kinnard, Aceto, and Buckley (1962) found that dosages of 1.5–3.0 mg/kg produced the attenuation of the CER, thereby perhaps providing some clue as to the widespread acceptance of the drug as effective in psychiatric settings (Gallant, Bishop, Steele, & Noblin, 1963; Hunter & Stephenson, 1962; Klein & Fink, 1963; Mann, 1963).

Heise and Boff (1962) have demonstrated that the conditioned avoidance model is superior for sensitivity to treatment effects over a number of other approaches (Holliday & Dille, 1958; Jindal, 1961; Randall, Heise, Schallek, Bagdon, Banziger, Boris, Moe, & Abrams, 1962; Weaver & Miya, 1961). Barry and Miller (1962), in an approach-avoidance study, found that 2–4 mg/kg reduced approach-avoidance speed in rats, and similar conclusions were reached by Harvey (1962). Barry (1962), in a well-controlled study, tended to corroborate these findings. Posluns (1962) found that 3 mg/kg produced a delay in acquisition of conditioned avoidance responses, as did Geller, Kulak, and Seifter (1962) with dosage levels of .25, .5, 1.0, and 3.0 mg/kg. Irwin (1962) found that the rate of tolerance in avoidance training was influenced by three variables: (*a*) females developed tolerance faster than males, (*b*) the intensity of the conditioned stimulus, (*c*) the intensity of the unconditioned stimulus. Experimenters have not typically equated these factors in avoidance studies; therefore, direct comparisons are not likely to permit any generalizations (Aston, Sekino, & Greifenstein, 1962; Ekstrom & Sandberg, 1962; Feldman & Lewis, 1962; Geller, Kulak, & Seifter, 1962; Jacob & Blozovski, 1961; Randall et al., 1961; Steward, 1962; Wolf et al., 1962).

One major problem in psychopharmacology concerns the effect of a given drug on an organism's discrimination of its own behavior, or the sensory feedback from this behavior (Berryman, Wagman, & Keller, 1960). These investigators, after establishing a stability criterion of performance for rats, injected them with .2 mg/kg ip 1 hour before the experimental session. The initial administration of the drug produced a noticeable deterioration of discriminative behavior, with subsequent administration the next day producing a much greater effect than the first. This study led the authors to conclude that chlorpromazine at a .2-mg/kg level would initiate a considerable change in discrimination against a previously established base line. The tendency for chlorpromazine to have successively greater effects for the same dosage if injected daily was substantiated by Olds and Travis (1960), using levels of 2 mg/kg with rats. With human subjects, Mirsky, Primac, and Bates (1959) found that 200 mg. of chlorpromazine produced significant impairment on the Continuous Performance Test of discriminant responding, and Shurtleff, Mostovsky, and DiMascio (1962), with dosages of 25, 50, 100, and 200 mg., found significantly impaired discrimination of tones at all frequencies.

Learning in Humans—Normals. A number of studies have been conducted which cast some light on the general effects of chlorpromazine on various measures of learning in both normal subjects and institutionalized mental patients. Studies on normals found generally impaired performance in reports ranging from the effects of chlorpromazine on a single bullfighter in the ring (Moraleda, 1961) to a multivariate evaluation of its effects on psychomotor and cognitive effects in a large number of subjects (Kornetsky, Humphries, & Evarts, 1957). The Kornetsky, Humphries, and Evarts study found increasing impairment on these tasks with increasing amounts of chlorpromazine. Kornetsky and

Humphries (1958) found that a single dose produced a greater decrement on psychomotor performance in normals than in schizophrenics. Schneider and Costiloe (1957) found it increased reaction time, and Burbridge (see Uhr & Miller, 1960) found it produced a decrement in key tapping and lowered the critical flicker fusion frequency. Similar decrements in CFF were found by Pollack, Karp, Krauthamer, Klein, and Fink (1961) and Karp and Pollack (1963).

Traugett, Balanov, and Kaufman (1958) found indications of impaired conditioned activity under chlorpromazine, and Schneider and Costiloe (1957) found that it inhibited reconditioning as well. Latz (1963), found that small dosage levels produced significant impairment in sustained attention, and that higher doses resulted in significantly impaired performance on attention, cognition, and psychomotor tasks. Although Ludvigson (1961) found no significant differences in conditioned eyelid responses between chlorpromazine (25 mg.) and placebo groups, he administered the drug orally 1 hour before testing, and there is reason to suspect that, in humans at least, a substantially longer time is required for results to be noticeable (Bennett & Hamilton, 1963; Clark, Ray, & Ragland, 1963b). A study showing no difference between chlorpromazine and placebo on reaction times and tapping rate was reported by Wenzel and Rutledge (1962). Mirsky and Cardon (1962) found a marked impairment on performance, and were able to validate their results (Mirsky et al., 1959). Using the same dosages used by Wenzel and Rutledge (1962); DiMascio, Havens, and Snell (1961), in a well-controlled study, found significant psychomotor inhibition, mental confusion, and sedation. Evans and Jewett (1962) found it increased disjunctive reaction time, and Schaacter and Wheeler (1959) found it inhibited humor. Uhr (1960) concluded that the majority of studies involving the stronger tranquilizers (e.g., chlorpromazine) impaired speeded, coordinated, psychomotor, and simple perceptual skills. Loomis and West (1958) attempted to assess the effects of chlorpromazine on a common activity and chose conditioned driving performance as their dependent variable. Using a simulated traffic condition, they found significant impairment in driving.

Psychiatric Patients. Results of research with chlorpromazine on hospitalized psychiatric patients show, for the most part, generalized decrements in a wide variety of learned behavior, and the few that show improvement typically do so at the expense of proper experimental controls. Ray, Ragland, and Clark (1962) gave Digit Symbol and Purdue Pegboard tests to patients during and after chlorpromazine treatment, and, finding no differences between these scores, concluded that the patients learned certain skills on drugs which were not lost when the drugs dissipated. This conclusion is not entirely tenable, however, since they ignored the base rate of predrug functioning level. In a later, well-controlled study (Clark, Ray, & Ragland, 1963a), they found stable significant differences, with earlier impairment on intellectual and psychomotor functions, but general effects requiring 8–16 weeks to become statistically significant. The long period of time that they felt was required to observe noticeable changes was also emphaized by Bennett and Hamilton (1963). A number of investigators have used the Wechsler-Bellevue scales to measure the effect of chlorpromazine on various subtests. Nickols (1962) reported a number of studies which indicated that placebo groups produced higher gains, especially on timed subtests, than chlorpromazine groups. Presumably, the gains occurred as practice effects due to familiarity with the (alternate form) test, and so might be considered a fair index of learning. Vestre (1961) found that chlorpromazine lowered both learning and retention, and Helper, Wilcott, & Garfield (1963) found similar results with emotionally disturbed children. Mitchell and Zax (1958) found that conditioning was impaired with chlorpromazine, and Saucer (1959) found that it raised the threshold for perception of apparent motion. Similar impairment of psychomotor and conditioning performance was reported by Kornetsky et al. (1959), and Porteus (1957a, 1957b) reported a decline in performance on maze tests.

Perhaps the ultimate criterion for the effect of drugs on learning capacity in psychiatric patients would be the effect of such agents on the adaptive capacity of medicated individuals in societal roles. Ellsworth and Clayton (1960) studied 200 successive psychotic patients admitted before drug therapy and 200 admitted after drug therapy. When matched for variables considered relevant, there were no statistically significant differences in either the length of hospitalization or the return rate of the two groups. Although the drug-treated patients left the hospital slightly sooner than those treated in the predrug area, they tended to return more rapidly than the predrug patients (31.3% of drug patients and 25.3% of predrug patients returned within one year). A year later Tuteur, Stiller, & Glotzer (1961) found that socioeconomic and other complicating factors interfered with the reintegration of the discharged patients being treated with chlorpromazine into society.

Aware of the problems involved in adjustment under drugs, Mason, Forrest, Forrest, and Butler (1963) emphasized the need for dosage adjustment and attention to psychosocial factors to help the patient meet the environmental stress of community living.

DISCUSSION

It is in the higher nervous system that human beings differ most from all other species of animal, so great caution must be stressed on making facile generalizations from infrahuman organisms to man. Many drug effects, such as those on individual differences, interpersonal relationships, and effects modifiable by experience, are essentially human variables which are difficult or impossible to isolate in lower species (Uhr & Miller, 1960). This may also be true, to a lesser extent, of the use of hospitalized psychiatric patients for behavioral testing. As Reitan (1960) has pointed out, there seems to be a large body of opinion in the field that normal subjects are better for such testing than patients, largely due to differential base-line considerations.

A statistical problem arises in the pooled testing of a number of variables in a single battery. With 20 such variables, chance factors alone should make one variable significant (.05) assuming that all variables are independent. Such a "significant" effect would have to be considered the result of chance, unless a replication were done for that single variable. This difficulty was graphically demonstrated in the study by Kelly, Miller, Marquis, Gerard, and Uhr (1958a, 1958b) in an extensive investigation of the effects of chronic administration of drugs on 51 behavioral measures with tests after each administration. They ran 153 tests for drug effects, plus 2,400 tests of significance of correlations between personality variables and drug effects and obtained a number of significant findings. But, as these were not clearly more than might have been expected by chance, it was concluded that no dependable drug effects had been found.

With a patient population there is a question concerning whether or not the individuals actually took their medication. Short of parenteral injection of the drug, perfect control is a questionable assumption to make. To sustain a constant level of the behavioral effect of a drug so that a battery of tests can be given, an investigator may either use a repeated injection method or spansules, although neither adequately insure the maintenance of a given drug level. As Newman (1941) pointed out, maintaining concentration of a drug at a uniform level in the blood, or even at the sites of drug action, does not in itself insure that the effects of the drug will be maintained at a uniform level. Further, the possibility of short-term adaptation of the central nervous system should be kept in mind when subjects are examined for longer periods of time (Beecher, 1952).

A number of rather pervasive difficulties make any definitive statements concerning the outcomes of the majority of the studies on the effects of drugs on learning somewhat tenuous. Few investigators have reported any behavioral base-line measures of learning prior to drug administration, and of these some have used base lines of atypical groups. There is rarely any control for such factors as institutionalization. When normals are compared with patients there is an almost universal tendency to ignore the fact that most patients are hospitalized whereas normals (by criterion) are usually not. It is perhaps a truism that few investigators in this field take into account the assumptions underlying various statistical tests. It is not uncommon to see 20 subjects put into cells of four each, with chi squares computed between and among all cells, for example, with a disregard for both the requirements for χ^2 and the effects of such multiple testing on the alpha level. A typical study was that of Denber, Teller, Rajotte, and Kaufman (1962) who used counterbalanced orders of 6 drugs and a placebo on a total sample of 38 patients.

A common assumption is that all "normals" are normal, without the substantiation of this fact empirically. It is extremely unusual for an examiner to ascertain that his normal subjects have no brain damage, even when comparing the effects of drugs on learning between normals and organics. Further, the "volunteer" subject, on whom many psychopharmacological studies with "normal" subjects depend, has been found to be rather atypical on a number of variables (Bellak, 1948; Hollister, Traub, & Beckman, 1956; Lasagna, 1955; Lindemann & Malamud, 1934; von Felsinger, Lasagna, & Beecher, 1955).

Although the effects of anxiety on learning performance are known (Spence, 1953), there appears to be little control for this variable other than in studies specifically investigating it (Holliday & Dille, 1958; Spence & Taylor, 1951, 1953). The improvements in conditionability reported by some investigators using lower dosage levels may be due in part to the alleviation of anxiety which impaired performance. Ferster et al. (1962), for example, reported increases in response rates with pigeons under lower and middle dosages, with a decrease at higher dosages. Similar results were reported by Schieckel (1962) with monkeys, and Kornetsky and Humphries (1958) found that under low dosages schizophrenics showed much less decrement than normals in various cognitive and psychomotor tasks.

It would appear that, in spite of the substantial numbers of studies investigating chlorpromazine, there has not been a conclusive, parametric investigation of the effects of this drug on learning, with controls for all variables found to be relevant. In the absence of such investigations no generalization to global situations is justified, but the nearly unanimous findings of the limiting effect of chlorpromazine on such a wide variety of learning measures does seem to support a conclusion that, since chlorpromazine is consensually considered the drug of choice for treatment of schizophrenic disorders (Freer, 1962), such treatment must be willing to sacrifice some degree of learning on the part of the patient.

REFERENCES

AARONSON, B. S. The comparative sensitivity of the qualitative and quantitative Porteus Maze scores to drug effects. *Journal of Clinical Psychology,* 1963, **19**, 184–185.

ACETO, M. D., LYNCH, V. D., & THOMS, R. K. Effects of drugs on conditioning in the rat: II. Synthesis of a centrally active drug and the effects of nine drugs on operant conditioning and extinction. *Journal of Pharmaceutical Sciences,* 1961, **50**(10), 823–827.

APPEL, J. B. Drugs, shock intensity, and the C. E. R. *Psychopharmacologia,* 1963, **4**, 148–153.

ASTON, R., SEKINO, E., & GREIFENSTEIN, F. E. Quantitation of drug effects upon conditioned avoidance behavior in the rat. *Toxicology and Applied Pharmacology,* 1962, **4**, 393–401.

BAKER, A. A., & THORPE, J. G. Placebo response. *Archives of Neurology and Psychiatry,* 1957, **78**, 57–60.

BARMACK, J. E. The effects of 10 mg. of benzedrine sulfate on the Otis test scores of college students. *American Journal of Psychiatry,* 1940, **97**, 163–166.

BARON, S. H., & FISHER, S. Psychotropic drug use in a general medical outpatient population. *Psychopharmacology Service Center Bulletin,* 1962. (Mimeo)

BARRY, H., III, & MILLER, E. Effects of drugs on approach-avoidance conflict tested repeatedly by means of a "telescope alley." *Journal of Comparative and Physiological Psychology,* 1962, **55**, 201–210.

BEECHER, H. K. Experimental pharmacology and measurement of the subjective response. *Science,* 1952, **116**, 157–162.

BEECHER, H. K., KEATS, A. S., MOSTELLER, F., & LASAGNA, L. The effectiveness of oral analgesics (morphine, codeine, acetylsalicylic acid) and the problem of placebo "reactors" and "non-reactors." *Journal of Pharmacology and Experimental Therapy,* 1953, **109**, 393–400.

BELLAK, L. *Dementia praecox.* New York: Grune & Stratton, 1948.

Bennett, I. F. Chemotherapy in psychiatric hospitals: Critical review of the literature and research trends. In, *Transactions of the first research conference on chemotherapy psychiatry*. Washington, D. C.: Veterans Administration, Department of Medicine and Surgery, 1957. Pp. 15–20.

Bennett, J. L., & Hamilton, L. D. Sequential use of antidepressant drugs with chlorpromazine in chronic schizophrenia. *Psychiatric Quarterly*, 1963, **37**, 53–66.

Berryman, R., Wagman, W., & Keller, F. S. Chlorpromazine and the discrimination of response-produced cues. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960. Pp. 243–249.

Bindra, D., & Mendelson, J. Training, drive level, and drug effects: A temporal analysis of their combined influence on behavior. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 183–189.

Boren, J. J. Some effects of adiphine, benactyzine, and chlorpromazine upon several operant behaviors. *Psychopharmacologia*, 1961, **2**, 416–424.

Clark, L. D. A comparative view of aggressive behavior. *American Journal of Psychiatry*, 1962, **119**, 336–341.

Clark, M. L., Ray, T., & Ragland, R. E. Chlorpromazine in chronic schizophrenic women: Rate of onset and rate of dissipation of drug effects. *Psychosomatic Medicine*, 1963, **25**, 212–217. (a)

Clark, M. L., Ray, T., & Ragland, R. E. Rate of onset and rate of dissipation of chlorpromazine effects and implications for the design of experiments. *Psychosomatic Medicine*, 1963, **24**, 520–521. (Abstract) (b)

Clark, R., Jackson, J. A., & Brady, J. V. Drug effects on lever positioning behavior. *Science*, 1962, **135**, 1132–1133.

Cole, Johanna O., & Gerard, R. W. (Eds.) *Psychopharmacology: Problems in evaluation*. (No. 583) Washington: National Academy of Sciences-National Research Council, 1959.

Cook, L., Kelleher, R. I., & Fellows, E. J. Pharmacodynamics of chlorpromazine and other phenothiazines. In J. H. Nodine & J. H. Moyer (Eds.), *Psychosomatic medicine*. Philadelphia: Lea & Febiger, 1962. Pp. 455–460.

Delay, J., Deniker, P., & Harl, J. M. The treatment of excitation and agitation states by a method of medication derived from hibernotherapy. *Annals of Method Psychology*, 1952, **110**, 267.

Denber, H. C., Teller, D. N., Rajotte, P., & Kaufman, Dorothy. Studies on mescaline XIII: A preliminary report. *Annals of the New York Academy of Sciences*, 1962, **96**, 13–36.

Dews, P. E., & Morse, W. Behavioral pharmacology. *Annual Review of Pharmacology*, 1961, **1**, 145–174.

DiMascio, A., Havens, L. L., & Snell, J. E. A comparison of four phenothiazine derivatives: A preliminary report on the assessment of chlorpromazine, promethazine, perphenazine, and trifluoperazine. In J. Wortis (Ed.), *Recent advances in biological psychiatry*, Vol. 3. New York: Plenum Press, 1961. Pp. 68–76.

Ekstrom, N. A., & Sandberg, F. A method for quantitative determination of the inhibitory effect on the conditioned avoidance response of mice. *Arzneimittel-Forschung*, 1962, **12**, 1208–1209.

Ellsworth, R. B., & Clayton, W. H. The effects of chemotherapy on length of stay, rate of return for psychiatrically hospitalized patients. *Journal of Consulting Psychology*, 1960, **24**, 50–53.

Estes, W. K., & Skinner, B. F. Some quantitative properties of anxiety. *Journal of Experimental Psychology*, 1941, **29**, 390–400.

Evans, W. O., & Jewett, A. The effect of some centrally acting drugs on disjunctive reaction time. *Psychopharmacolgia*, 1962, **3**, 124–127.

Everett, G. M. Some electrophysiological and biochemical correlates of motor activity and aggressive behavior. In, *Neuro-psychopharcology*. Vol. 2. Amsterdam: Elsevier, 1961. Pp. 479–484.

Feer, H. Comparison of the clinical effects of three phenothiazine derivatives. In R. A. Cleghorn, A. E. Moll, & C. A. Roberts (Eds.), *Third world congress of psychiatry proceedings*. Vol. 2. Toronto: Univer. Toronto Press, 1962. Pp. 964–966.

Feldman, R. S., & Lewis, E. Response differences of psychotropic drugs in rats during chronic anxiety state. *Journal of Neuro-psychiatry*, 1962, **3**, 527–541.

Ferster, C. R., Appel, J. B., & Hiss, R. A. The effects of drugs on a fixed-ratio performance suppressed by a pre-time-out stimulus. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 73–88.

Finocchio, D. V. Drug effects on "emotionally" disrupted timing behavior. *Federation Proceedings*, 1963, **22**, 511. (Abstract)

Fleischmann, E. A. A factor analysis of intra-task performance on two psychomotor tasks. *Psychometrika*, 1953, **18**, 45–55.

Fleischmann, E. A., & Hempel, W. E., Jr. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, **19**, 239–252.

Fleischmann, E. A., & Hempel, W. E., Jr. The relationship between abilities and improvement with practice in a visual discrimination reaction task. *Journal of Experimental Psychology*, 1955, **49**, 301–312.

Freyhan, F. A., & Merkel, J. Clinical and social aspects of compensatory drug treatment. In M. Greenblatt, D. J. Levinson, & C. L. Kleiman (Eds.), *Mental patients in transition*. Springfield, Ill.: Charles C Thomas, 1961. Pp. 302–312.

Gallant, D. M., Bishop, M. P., Steele, C. A., & Noblin, D. D. The relationship between sorontin antagonism and tranquilizing activity. *American Journal of Psychiatry*, 1963, **119**, 882.

Gatti, G. L., & Frank, M. The technique of the "pole climbing" response. Effect of various types of drugs on the latency period and flight reaction of the rat. In, *Neuro-psychopharmacology*. Vol. 2. Amsterdam: Elsevier, 1961. Pp. 147–150.

Geller, I., Kulak, J. T., Jr., & Seifter, J. The effects of chlorpromazine on a punishment discrimination. *Psychopharmacologia*, 1962, **3**, 374–385.

Gonzalez, R. C., & Shepp, B. The effects of chlorpromazine on instrumental learning based on conditional fear. *Canadian Journal of Psychology*, 1962, **16**, 64–71.

Halasz, M. The extinction of intercurrent avoidance and appetitive responses under chlorpromazine. Unpublished doctoral dissertation, University of Chicago, 1959.

Harvey, J. Behavioral effects of pharmacological agents. *American Psychologist*, 1962, **17**, 397.

Heimstra, N. W. Social influence on the response to drugs: II. Chlorpromazine and iproniazid. *Psychopharmacologia*, 1962, **3**, 72–78.

Heise, G. A., & Boff, E. Continuous avoidance as a baseline for measuring behavioral effects of drugs. *Psychopharmacologia*, 1962, **3**, 264–282.

Helper, M. M., Wilcott, R. C., & Garfield, S. L. Effects of chlorpromazine on learning and related processes in emotionally disturbed children. *Journal of Consulting Psychology*, 1963, **27**, 1–9.

Holliday, Audrey R., & Dille, J. M. The effects of meprobamate, chlorpromazine, pentobarbital, and a placebo on a behavioral task under stress conditions. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 811–815.

Hollister, L. E., Traub, L., & Beckman, W. G. Psychiatric uses of reserpine or chlorpromazine: Results of double-blind studies. In N. S. Kline (Ed.), *Psychopharmacology: A symposium organized by the Section in Medical Sciences of the AAAS and the APA presented at the Berkeley meeting December 30, 1954*. Washington, D. C.: American Association for the Advancement of Science, 1956. Pp. 65–74.

Hunt, H. F. Some effects of drugs on classical (type S) conditioning. *Annals of the New York Academy of Science*, 1956, **65**, 258–267.

Hunt, H. F. Some applications of conditioned "fear." In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960. Pp. 227–232.

Hunter, H., & Stephenson, G. M. Chlorpromazine and trifluoperazine in the treatment of behavioral abnormalities in the severely subnormal child. *British Journal of Psychiatry*, 1963, **109**, 411–417.

Irwin, S. Influence of experimental variables on the rate of tolerance development to chlorpromazine and perphenazine. *Federation Proceedings*, 1962, **21**, 419. (Abstract)

Irwin, S., & Armstrong, Patricia. Conditional locomotor response with drug as the unconditioned stimulus: Individual differences. *Neuropsychopharmacology*, 1961, **2**, 151–157.

Izquierdo, I. Observations on conditioned and unconditioned on-and-off behavioral responses to a buzzer. *Journal of Pharmacy and Pharmacology*, 1962, **14**, 316–318.

Jacob, J., & Blozovski, M. Effects of various analgesic agents on the behavior of mice subjected to a thermoalgesic stimulus: II. Learning under nocioceptive stress conditions. Comparative effects of analgesic and psychoactive agents on linking and leaping reactions. *Archives Internationales de Pharmacodynamie et de Thérapie*, 1961, **133**(3-4), 296–309. (Abstract)

Jellinek, E. M. Clinical tests on comparative effectiveness of analgesic drugs. *Biometrics Bulletin*, 1946, **2**, 87–91.

Jellinek, E. M., & McFarland, R. A. Analysis of psychological experiments on the effects of alcohol. *Quarterly Journal of Studies on Alcohol*, 1940, **1**, 272–371.

Jernberg, P. The effects of chlorpromazine upon the acquisition and loss of a conditioned emotional response. Unpublished master's thesis, University of Chicago, 1957.

Jindal, M. N. Neuromuscular blockade by some phenothiazine derivatives. *Archives Internationales de Pharmacodynamie et de Therapie*, 1961, **132** (3-4), 322–330.

Karp, E., & Pollack, M. Comparative studies of chlorpromazine and imipramine: III. Critical flicker fusion thresholds in psychiatric patients. *Psychopharmacologia*, 1963, **4**, 452–458.

Kelly, E. L., Miller, J. G., Marquis, D. G., Gerard, R. W., & Uhr, L. Continued meprobamate and prochlorperazine administration and behavior. *Archives of Neurology and Psychiatry*, 1958, **80**, 247–253. (a)

Kelly, E. L., Miller, J. G., Marquis, D. G., Gerard, R. W., & Uhr, L. Personality differences and continued meprobamate and prochlorperazine administration. *Archives of Neurology and Psychiatry*, 1958, **80**, 241–246. (b)

Kinnard, W. J., Aceto, M. D. G., & Buckley, J. P. The effects of certain psychotropic agents on the conditioned emotional response behavior of the albino rat. *Psychopharmacologia*, 1962, **3**, 227–230.

Klein, D. F., & Fink, M. Multiple item factors as change measures in psychopharmacology. *Pharmacologia*, 1963, **4**, 43–52.

Kornetsky, C., & Humphries, O. Psychological effects of centrally acting drugs in man. Effects of chlorpromazine and secobarbital on visual and motor behavior. *Journal of Mental Science*, 1958, **104**, 1093–1099.

Kornetsky, C., Humphries, O., & Evarts, E. V. Comparison of psychological effects of certain centrally acting drugs in man. *Archives of Neurology and Psychiatry*, 1957, **77**, 318–324.

Kornetsky, C., Pettit, M., Wynne, R., & Evarts, E. V. A comparison of the psychological effects of chlorpromazine and secobarbital in schizophrenics. *Journal of Mental Science*, 1959, **105**, 190–199.

Kovitz, B., Carter, J. T., & Addison, W. P. A comparison of chlorpromazine and reserpine in chronic psychoses. *Archives of Neurology and Psychiatry*, 1955, **74**, 467–471.

Kris, Else B. Children born to mothers maintained on pharmacotherapy during pregnancy and postpartum. In J. Wortis (Ed.), *Recent advances in biological psychiatry*. Vol. 4. New York: Plenum Press, 1962. Pp. 180–187.

Lasagna, L. The controlled clinical trial: Theory and practice. *Journal of Chronic Diseases,* 1955, **1**, 353–367.

Lasagna, L., Mosteller, F., von Felsinger, J. M., & Beecher, H. K. A study of the placebo response. *American Journal of Medicine,* 1954, **16**, 770–779.

Latz, A. The differential effects of chlorpromazine and secobarbital on sustained attention in chronic schizophrenics. *Federation Proceedings,* 1963, **22**, 509. (Abstract)

Lindemann, E., & Malamud, W. Experimental analysis of the psychopathological effects of the intoxicating drugs. *American Journal of Psychiatry,* 1934, **90**, 853–879.

Loomis, T. A., & West, T. C. Comparative sedative effects of a barbiturate and some tranquilizer drugs on normals. *Journal of Pharmacology,* 1958, **122**, 525–531.

Ludvigson, H. W. *The effects of chlorpromazine and d-amphetamine on eyelid conditioning.* (Doctoral dissertation, State University of Iowa) Ann Arbor, Mich.: University Microfilms, 1961, No. 60–5678.

Mann, J. The initiation of treatment of unmanageable psychotics with intramuscular administration of trifluoperazine combined with chlorpromazine. *American Journal of Psychiatry,* 1963, **120**, 74–76.

Mason, A. S., Forrest, Irene, Forrest, F. M., & Butler, H. Adherence to maintenance therapy and rehospitalization. *Diseases of the Nervous System,* 1963, **24**, 103–104.

McNemar, Q. On the use of latin squares in psychology. *Psychological Bulletin,* 1961, **48**, 398–401.

Mirsky, A. F., & Cardon, P. V. A comparison of the behavioral and physiological changes accompanying sleep deprivation and chlorpromazine administration in man. *Encephalography and Clinical Neurophysiology,* 1962, **14**, 1–10.

Mirsky, A. F., Primac, D. W., & Bates, R. The effects of chlorpromazine and secobarbital on the continuous performance test. *Journal of Nervous and Mental Disorders,* 1959, **128**, 12–17.

Mitchell, L. E., & Zax, M. The effects of ataractic therapy on conditioning rate of G. S. R. in a group of psychiatric patients. *American Psychologist,* 1958, **13**, 343. (Abstract)

Moraleda, A. Studies on the effects of several tranquilizing drugs on a professional bullfighter during the course of the season. *Actas-Luso-Espanolas de Neurologia y Psiquiatria.* 1961, **20**, 239–241. (Abstract)

Munroe, Ruth. *Schools of psychoanalytic thought.* New York: Holt, Rinehart, & Winston, 1955.

Murphree, O. D., Monroe, B. L., & Seager, L. D. Survival of offspring of rats administered phenothiazines during pregnancy. *Journal of Neuropsychiatry,* 1962, **3**, 295–297.

Nash, H. The design and conduct of experiments. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley, 1960. Pp. 128–155.

Newman, H. W. *Acute alcoholic intoxication.* Palo Alto, Calif.: Stanford Univer. Press, 1941.

Nickols, J. Tranquilizers and experimental psychopathology. *American Psychologist,* 1962, **5**, 263–264.

Olds, J., & Travis, R. P. Effects of drugs on self stimulation. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley, 1960. Pp. 255–268.

Ostow, M. *Drugs in psychoanalysis and psychotherapy.* New York: Basic Books, 1962.

Paillard, J., Bert, J., Zwingelstein, J., & Gindicelli, P. Recherche d'une méthode d'approche de l'action neurophysiologique de diverses droques psychotropes chez l'homme. *Revue Neurologique,* 1961, **3**, 227–228. (Abstract)

Paulter, E. L., & Clark, G. The effect of chlorpromazine on the discrimination between intermittent photic stimulation and a steady light in normal and brain-damaged cats. *Journal of Comparative and Physiological Psychology,* 1961, **54**-(5), 493–497.

Plotnikoff, N. Drug resistance due to inbreeding. *Science,* 1961, **134**, 1881–1882.

Pollack, M., Karp, E., Krauthamer, G., Klein, D. F., & Fink, M. Neuropsychologic response patterns of some psychotropic drugs. In, *Neuropsychopharmacology.* Vol. 2. Amsterdam: Elsevier, 1961. Pp. 381–384.

Porteus, S. D. Maze test reactions after chlorpromazine. *Journal of Consulting Psychology,* 1957, **21**, 15–21. (a)

Porteus, S. D. Specific behavior changes following chlorpromazine. *Journal of Consulting Psychology,* 1957, **21**, 257–263. (b)

Posluns, D. An analysis of chlorpromazine-induced suppression of the avoidance response. *Psychopharmacologia,* 1962, **3**, 361–373.

Rahmann, H. The influence of chlorpromazine on memory in the golden hamster. *Pflüger's Archives,* 1962, **275**, 95–106. (Abstract)

Randall, L. O., Heise, G. A., Schallek, W., Bagdon, R. E., Banziger, R., Boris, A., Moe, R. A., & Abrams, W. B. A new psychotherapeutic agent of the benzodiazepine class. *Current Therapeutic Research,* 1961, **3**(9), 405–425.

Ray, T. S., Ragland, R. E., & Clark, M. L. Onset and decline of chlorpromazine effects on various psychological modalities in chronic schizophrenic women. *Federation Proceedings,* 1962, **21**, 419. (Abstract)

Reitan, R. M. Comparative behavioral effects of phenagylycose, meprobamate and placebo. In L. Uhr and J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley, 1960. Pp. 365–368.

Rosenberg, F. J., & Savaris, P. J. Reversal of chlorpromazine depression. *Federation Proceedings,* 1963, **22**, 568. (Abstract)

Rosenthal, D., & Frank, J. D. Psychotherapy and the placebo effect. *Psychological Bulletin,* 1956, **53**, 294–302.

Russell, R. W. Psychopharmacology. *Annual Review of Psychology,* 1964, **15**, 94–114.

SADOWSKI, B., & LONGO, V. G. Electroencephalographic and behavioral correlates of an instrumental reward conditioned response in rabbits. A physiological and pharmacological study. *Electroencephalography and Clinical Neurophysiology,* 1962, **14**, 465–476.

SAUCER, R. T. Chlorpromazine and apparent motion perception by schizophrenics. *Journal of Consulting Psychology,* 1959, **23**, 134–136.

SCHACHTER, S., & WHEELER, L. Epinephrine, chlorpromazine, and amusement. *Journal of Abnormal and Social Psychology,* 1962, **65**, 121–128.

SCHIECKEL, C. L. Effects of chlorpromazine and chlordiazepoxide on delayed matching in the monkey measured by a self-adjusting procedure. *American Psychologist,* 1962, **17**, 398.

SCHNEIDER, R. A., & COSTILOE, J. P. Effects of centrally acting drugs on conditioning man; the inhibiting and facilitating effects of chlorpromazine, amobarbital, and methyephenidylacetate on the conditioned galvanic skin response. *American Journal of Medical Science,* 1957, **233**, 418–423.

SHATEN, L., ROCKMORE, L., & FUNK, I. C. Responses of psychiatric patients to massive doses of thorazine: II. Psychological test performance and comparative drug evaluation. *Psychiatric Quarterly,* 1956, **30**, 402–416.

SHURTLEFF, D., MOSTOFSKY, D., & DIMASCIO, A. The effects of some phenothiazine derivatives on the discrimination of auditory clicks. *Psychopharmacologia,* 1962, **3**, 153–165.

SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In, *Learning theory, personality theory, and clinical research.* New York: Wiley, 1953, Pp. 1–21.

SPENCE, K. W., & TAYLOR, JANET. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *Journal of Experimental Psychology,* 1951, **42**, 183–188.

SPENCE, K. W., & TAYLOR, JANET. The relation of conditioned response strength to anxiety in normal, neurotic and psychotic subjects. *Journal of Experimental Psychology,* 1953, **45**, 265–272.

STEWARD, JANE. Differential responses based on the physiological agents. *Psychopharmacologia,* 1962, **3**, 132–138.

TRAUGETT, N. N., BOLVANOV, L., & KAUFMAN, D. A. Mechanism of the action of chloropromazine on the higher nervous activity of man. *Zhurnal vÿsshei nervnoĭ deyatel'nosti imeni I. P. Pavlova,* 1958, **11**, 814–822. (Abstract)

TRAVIS, R. P., & OLDS, J. Effects of secobarbital on escape and avoidance behavior. *American Psychologist,* 1959, **14**, 430.

TUTEUR, W., STILLER, R., & GLOTZER, J. Chlorpromazine—five years later. In J. Wortis (Ed.), *Recent advances in biological psychiatry.* Vol. 3. New York: Grune & Stratton, 1961. Pp. 35–43.

UHR, L. Objectively measured drug effects. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley, 1960. P. 614.

UHR, L., & MILLER, J. G. (Eds.) *Drugs and behavior.* New York: Wiley, 1960.

VESTRE, N. D. *An experimental study of the effects of chlorpromazine on association learning in schizophrenics.* (Doctoral dissertation, University of Minnesota) Ann Arbor, Mich.: University Microfilms, 1961, No. 60-5623.

VON FELSINGER, J. M., LASAGNA, L., & BEECHER, H. K. Drug-induced mood changes in man. 2. Personality and reactions to drugs. *Journal of the American Medical Association,* 1955, **157**, 1113–1119.

WEAVER, M. E., & MIYA, T. S. Effects of certain ataraxic agents on mice activity. *Journal of the Pharmaceutical Sciences,* 1961, **50**(11), 910–912.

WEITZMAN, E. D., & ROSS, G. S. A behavioral method for the study of pain perception in the monkey. *Neurology,* 1962, **12**, 264–272.

WENZEL, D. G., & RUTLEDGE, C. O. Effects of centrally-acting drugs on human motor and psychomotor performance. *Journal of Pharmaceutical Sciences,* 1962, **51**, 631–644.

WERBOFF, J., & KESNER, R. Learning deficits of offspring after administration of tranquilizing drugs to the mothers. *Nature,* 1963, **197**, 106–107.

WHITE, R. P., & WESTERBEKE, E. J. Differences in central anticholinergic actions of phenothiazine derivatives. *Experimental Neurology,* 1961, **4**(4), 317–329.

WOLF, H. H., SWINYARD, E. A., & CLARK, L. D. The differential effects of chlorpromazine and pentobarbital on two forms of conditioned avoidance behavior in *Peromyscus maniculatus gracilis. Psychopharmacologia,* 1962, **3**, 438–448.

WOLPE, J., SALTER, A., & REYNA, L. J. (Eds.) *The conditioning therapies.* New York: Holt, Rinehart, & Winston, 1964.

(Received March 1, 1965)

Psychological Bulletin
1965, Vol. 64, No. 4, 246–258

VISION IN THE GANZFELD

LLOYD L. AVANT[1]

Kansas State University

A summary of the evidence on exposure to structureless visual fields is presented. The data show the experience of such fields to be characterized by reports of: immersion in a "sea of light" which separates into figure and ground as brightness is increased, chromatic adaptation in colored fields, loss of efficiency in detecting the presence and movement of inhomogeneities introduced into the field, disorientation of the observer, an increased and fluctuating state of accommodation, and the occasional joint occurrence of an apparent cessation of function of the visual mechanism and increased alpha activity in the brain.

The advent of high-altitude supersonic flight and space travel have added to the theoretical concern for the Ganzfeld a practical interest in the effects of a structureless field upon vision. The loss in efficiency of the visual mechanism in the Ganzfeld is similar to that of high-altitude flight. This similarity is exemplified by the fluctuation and increase in accommodation characteristic of "empty field" myopia for the Ganzfeld observer and "sky" myopia in the case of high altitude flying. The purpose of this paper is to review and discuss available evidence on the effects of exposure to structureless visual fields.

In 1930 Metzger seated observers 1.25 meters distant from a carefully whitewashed 4 × 4 meter square surface from which wings extended toward the observer on three sides. He illuminated the field with a neutral light and asked the observers to verbalize their experience of this field. Metzger termed this field the Ganzfeld. Gestalt theory, holding that primitive vision is binocular rather than monocular, would predict for Metzger's observers an experience of tridimensional, rather than bidimensional, space. Metzger's data favored this Gestalt prediction, and the contribution of Metzger's work to Gestalt theory is reflected in Koffka's (1935) statement of Gestalt thinking.

Metzger's results did not, however, completely support the Gestalt position, and Metzger's query, as well as his new method of studying visual perception, went unnoticed for some years. The revival of interest in the experience of the Ganzfeld has occurred largely during the last decade. More recent research has shown that, with regard to a number of visual functions or skills, the Ganzfeld produces an experience markedly different from the experience of structured fields. Varying the brightness and/or chromaticity of the field or the introduction of spots differing in brightness or chromaticity from the remainder of the field produces responses to such variations which differ from responses to these variables in structured fields. The observer's accommodation and orientation are affected in the Ganzfeld situation, and the visual mechanism even appears at times to cease functioning without field structure.

PHENOMENAL CHARACTERISTICS OF THE GANZFELD

Illuminating the Ganzfeld with neutral light, Metzger (1930; cited by Koffka, 1935) found that the observer felt "himself swimming in a mist of light which becomes more condensed at an indefinite distance." Practically all subsequent investigators have reported a like experience for the observer in the Ganzfeld. Impressions reported by Hochberg, Triebel, and Seaman (1951) were such as to be subsumed under the term "fog" or "cloud." According to Gibson and Waddell (1952) the experience was that of a "sea of light," with some observers at times sensing something vaguely surfacelike in front of the face. Cohen's (1956, 1957) observers found it difficult to apply to the empty field

---

[1] The author wishes to thank Harry Helson and William Bevan for their critical reading of this paper.

experience the language usually adequate to express visual experience of structured fields; "sea of light" seemed most descriptive for most of these observers.

Atypical experience includes surfaces and a "cone-shaped three-dimensional surface" concave toward the observer (Hochberg, Triebel, & Seaman, 1951), a "cracked ice effect" or a "web-like structure" (Cohen, 1957), and awareness of vitreous opacities and retinal blood vessels (Miller & Hall, 1962). Miller and Ludvigh (1960) reported gradations of light and shade, rings of different sizes and shapes, and when homogeneous stimulation of other sense modalities was added to the experience of the Ganzfeld, extensive hallucinations resulted (Bexton, Heron, & Scott, 1954).

The location, or distance, of the fog produced by the Ganzfeld is not accurately determined. Cohen's (1957) observers described the homogeneous field as "close at hand," although the meaning of "close" varied among observers. Only one of these observers consistently described the field as being more than 6 inches away, and the modal distance was 2 inches. However, the fog extended for an indefinite distance. Gibson and Waddell (1952) pointed out the voluminous nature of the experience of the fog. Their report is in agreement with other investigators and is the basis for Gibson and Waddell's tentative agreement with Metzger's findings in support of the Gestalt belief in the tridimensional nature of primitive visual experience. However, Gibson and Waddell stressed caution in such a conclusion because observers agree that the experience is indefinite, indeterminate, or ambiguous. Neither Metzger nor Gibson and Waddell have given definitive evidence of the two-dimensional or three-dimensional nature of primitive visual experience.

Various aftereffects of exposure to homogeneous stimulation for a period of 20 minutes were found in one study (Cohen, 1960). The observers of this study experienced extreme fatigue and a feeling of great lightness of body. Motor coordination was reportedly poor, and observers had difficulty maintaining balance. Time perception was disturbed. Subjects often complained of dizziness and sometimes appeared to be intoxicated. One observer experienced temporary states of depersonalization, a condition which Cohen did not define.

BRIGHTNESS VARIATIONS IN THE GANZFELD

The "mist of light" experience described by Metzger's observers prevailed as long as the illumination of the whitewashed surface remained "below a certain level." Increasing the illumination produced new experiences. As the illumination was increased

> the fog became condensed into a regularly curved surface which surrounded the observer on all sides; its appearance was filmy like the sky, not surfacy, and, similar to the sky, it was slightly flat in the center [Metzger, 1930, p. 13; cited by Koffka, 1935].

Further increase in the illumination caused the surface to straighten out into a plane, the apparent distance of which was definitely greater than that of the real one. The explanation of this result offered by Koffka was that the increased illumination allowed the observer to accommodate to the microstructure or "grain" of the surface, and there was no longer homogeneous stimulation. Under Metzger's conditions, these results are not surprising. It is of interest to note, however, that Metzger's results tend to support Ittelson and Ames' (1950) contention that accommodation is to apparent rather than to objective distance. Metzger's observers, given sufficient illumination to accommodate to the microstructure of the viewed surface, saw it as farther away than was actually the case. Ittelson and Ames varied the apparent distance of playing cards by changing the size of the cards; they found, with a constant distance for cards of different sizes, that accommodation and convergence varied with apparent distance.

Cohen (1956, 1957, 1958a, 1958b) produced the Ganzfeld differently from Metzger and obtained different indications of the effects of brightness variations. His Ganzfeld was produced by two adjoining spheres, A and B, 1 meter in diameter, which, at their juncture, produced an 8-centimeter aperture. Wearing a special mask which also formed a removable section of the wall of Sphere A, these observers looked monocularly into Sphere A. By controlling the chromaticity, purity, and luminance of the two spheres independently, the experimenter could produce a homogeneous field or a field containing a background (Sphere A) and an 8-centimeter spot (the juncture of Spheres A and B). In the homogeneous field Cohen found that variations in the intensity of the field did not consistently alter the appearance of the field. There was, however, some tendency for the fog of brighter fields to appear denser and closer.

With a field made inhomogeneous by the introduction of the spot, he found that variations in brightness differences between field and spot within the range 1.8 to 4.9 millilamberts produced no statistical evidence of a relation between intensity of illumination and the density or closeness of the fog. This lack of agreement with Metzger's findings Cohen attributes to the microstructure of Metzger's field which was absent in his own. With regard to the separation of field and spot, Cohen found that when the gradient between field and spot is not steep the fog seems nearer but the spot seems farther away. Further increase in the intensity gradient between field and spot reverses this appearance; the fog recedes, and the spot is seen in front of the field fog. At a maximum intensity difference between field and spot, judgments of *figural* distance ranged from "next to the eye" to 3 feet away with a median of 6 inches. The ground was usually judged a few inches behind the spot. Cohen reports great individual differences between, but consistency within, distance judgments of his observers. In addition to these effects of independent variation of brightness, this investigator found that when inhomogeneity was initially produced by chromatic differences the figure-ground separation thus established was enhanced by the addition of a brightness gradient. Cohen concluded that the phenomenal characteristics of the Ganzfeld are not independently determined. He maintained that any change in the stimulus distribution that modifies one aspect of experience produces concomitant changes in other phenomenal characteristics. of the Ganzfeld. The most recent study of the Ganzfeld (Weintraub, 1964) indicates that Cohen's conclusion is incorrect, and that intensity of stimulation and the colorimetric purity of the Ganzfeld are independent dimensions. Weintraub's results appear to disagree with the well-established principle of color conversion (Helson, 1938), certainly with regard to the chromatic effect of brightness variation. This conflict is discussed in the following section.

CHROMATIC EFFECTS IN THE GANZFELD

Questioning the necessity of the experience of a mid-gray as the end result of prolonged homogeneous stimulation, as required by Hering's opponent processes theory of vision, Helson and Judd (1932) constructed an "adaptation sphere" of Hering paper, suspended a 100-watt Mazda lamp in its center, and had observers sit with the sphere encircling their heads for as long as 75 minutes. They report that whatever chromatic adaptation took place occurred during the first few minutes, but at no time, even after 75 minutes, did the color of the surface of the sphere disappear. The paper of which the inner surface of the sphere was constructed was an orange-red, and the authors state that the red component tended to be reduced, making the orange surface more yellowish. Only during very brief periods lasting no more than a few seconds when the eyes were held fixed did a fairly good gray appear. These authors interpret their results as evidence against the Hering theory of vision with its concept of a mid-gray as the end result of adaptation. They concluded that "eye-movements alone, all other factors held constant, are sufficient to arrest the course of adaptation so that complete adaptation with total loss of chroma does not occur [p. 397]." Hochberg, Triebel, and Seaman (1951) criticized Helson and Judd's results on the basis that they were obtained in a field of definite structure. However, Helson [2] reported an unpublished confirmation of his and Judd's findings and pointed out (Helson, 1964) that complete adaptation is not found with very intense stimulation.

The results obtained by Hochberg, Triebel, and Seaman (1951) concerning chromatic effects in the Ganzfeld contrast sharply with those obtained by Helson and Judd

---

[2] Personal communication, 1965.

(1932). In the first of a series of studies, Hochberg, Triebel, and Seaman produced a homogeneous field by taping halved table-tennis balls over the eyes of their observers; they stimulated one group by a 100-watt light passed through a red Wratten filter (no. 70) and another by the same light passed through a green Wratten filter (no. 54). In the group stimulated by red light, 5 out of 6 observers initially reported a red-colored surfaceless field followed by the disappearance of the color within the first 3 minutes. The remaining observer of this group reported a three-dimensional surface which remained a "very dark, brownish magenta." All the observers of the group stimulated with green light initially reported a green surfaceless field followed within a maximum of 6 minutes by total disappearance of the green. The one point on which the Helson and Judd and the Hochberg, Triebel, and Seaman studies agree is that the final state of chromatic adaptation is not the mid-gray postulated by Hering. Whereas Helson and Judd reported little color loss, Hochberg, Triebel, and Seaman reported final adaptation to range from black (or "nothingness") to dark gray.

In further phases of their work, Hochberg, Triebel, and Seaman covered the right eye while the left was adapted to either a red or a green light. They stated that when the covering was removed from the right eye the observer reported the perception of the illuminant color and three fifths of their observers localized this perception in the non-adapted right eye. Again, they reported the loss of color perception with the right eye after prolonged stimulation, although this adaptation was not as satisfactory as that obtained with simultaneous stimulation of both eyes or of one eye alone. The movement of a finger so as to cast a shadow on the lower portion of the field of the adapted eye produced the experience of a black shadow with a halo of the color of the illuminant. Eye movements, after the loss of color was reported by the observer, produced no reappearance of color for half the observers and fleeting flashes of either red or blue-green for the other half. Accordingly, the authors concluded that eye movements have only limited ability to restore color perception. Interruption of the red light for approximately 2 seconds produced a complementary blue-green for 9 out of these 10 observers, and reestablishment of the red illuminant produced a brief flash of red.

Hochberg, Triebel, and Seaman's use of halved table-tennis balls to produce homogeneous stimulation deserves notice. Other investigators have used this technique (e.g., Weintraub, 1964). It is highly likely that this technique produces a different field from that produced by a larger stimulus field at a greater distance from the eyes.

Producing the homogeneous field by a technique different from that of either Metzger, Helson, and Judd, or Hochberg, Triebel, and Seaman, Cohen (1956, 1957, 1958b) also found a loss of chroma within an adaptation period of 3 minutes for 80% of his observers, the remaining 20% reporting only a "trace of color." As previously noted, Cohen produced either a homogeneous or inhomogeneous field by varying the chroma and/or luminance of two adjoining spheres. The observer looked into one of these spheres, and both the luminance and chromaticity of each sphere were independently controlled. The observer's visual field could be made either homogeneous or inhomogeneous by the presence or absence of color or brightness gradients between the two spheres.

The reports from Cohen's (1957, 1958b) observers are largely in agreement with the results obtained by Hochberg, Triebel, and Seaman (1951). The most noticeable difference in the results of these studies is that Cohen's results would support the medium gray end state of color adaptation called for by the Hering theory, whereas Hochberg, Treibel, and Seaman reported an end state of too dark a gray to be in accord with Hering's position. Cohen suggested that this difference may be attributable to the fact that Hochberg, Triebel, and Seaman used a lower intensity of illumination and longer periods of adaptation. Cohen also reported that dominant, long wavelength illumination made the Ganzfeld fog appear closer and denser than did dominant, short wavelength illumination. His observers reported feeling less "immersed" in the "blue" field, and to Cohen this suggested that the receptors sensitive to long wavelengths are more densely concentrated than those sensitive to short wavelengths.

Cohen's introduction of a differentiated area into the field changed its phenomenal characteristics. When this differentiation was due only to chromatic differences between the field and the 8-centimeter spot, the field was described as more saturated than was the homogeneous field. The addition of an intensity difference to the chromatic difference reduced the saturation of the field to the level reported for the homogeneous field. This was true except when the difference in intensity was such as to make the spot achromatic and black; in this case, the saturation of the field surrounding the spot was increased. And, as might be expected by classical contrast, most of the saturation enhancement occurred in the area of the spot figure. Cohen (1958b) stated that the presence of a *chromatic* difference within the field does not radically change most of the phenomenal characteristics of the Ganzfeld, that its principal effect is on the experience of hue. It is noteworthy, however, that a chromatic difference produces a separation of figure and ground, albeit a relatively indefinite one. With a chromatic difference, observers described a well-saturated figure in a poorly saturated ground, even when the stimulation from the spot was achromatic. Again, Cohen pointed out that this separation is primarily in one dimension of experience only, that is, color. With larger intensity gradients between figure and field, separation of the two becomes possible on other dimensions as well; such aspects as spatial separation, differences of contour, and textural differences between figure and ground become more pronounced, and a reduction of saturation in the field occurs. As noted above, with moderate intensity gradients Cohen's observers reported the fog of the field nearer and the differentiated figure farther away, while an increased intensity gradient reversed this relation, causing the field fog to recede and the figure to approach the observer.

The most characteristic effect of adaptation to inhomogeneity in the field Cohen (1958b) found to be the formation of a halo, usually involving desaturation and darkening of the field, immediately following the formation of a small, bright, highly saturated area encircling the spot. This halo may consist of a mixture of colors which are not fused, but appear to exist independently in different planes. Although spatial characteristics of the halo vary, it was most often seen as being in a plane behind the figure but in front of the field, appearing, thus, to facilitate the separation of figure and ground.

Although it was performed in a structured field, a study reported by Helson and Jeffers (1940) provides an interesting comparison with Cohen's findings in a chromatic field. Cohen reported the effects of chromatic inhomogeneity to be restricted largely to the color dimension of experience. Helson and Jeffers found that object colors lose their individuality to a great extent when the illuminant changes from daylight to strongly chromatic. With such a change they said that colors retreat into the plane of the background, that they become softer and filmier, that demarcation between object and background colors is not sharp, and contours almost disappear. In addition, the observer seems to be bathed in the illumination which takes on a strong affective character. Chromatic illumination results in perceptions approaching those of Ganzfeld conditions.

In two additional studies, cessation of the experience of color after prolonged exposure to an empty field was found. An abstract of a study unavailable to this writer by ten Doesschate (1961) stated that, "When the empty field is viewed through colored glasses, the perception of chroma gradually diminishes and is often followed by a sensation of dark grey. . . . [and in two cases] periods of apparent darkness [p. 565]."

The most recent study conducted under empty field conditions (Weintraub, 1964) produced results apparently conflicting with the well-established principle of color conversion (Helson, 1938). Helson's statement of this principle is that in every viewing situation an adaptation level (AL) is established such that objects having *reflectances* above the AL will be tinged with the hue of the illuminant, objects having reflectances below the AL will be tinged with the hue of the afterimage complement of the illuminant, and objects having reflectances at or near the AL will be either achromatic or weakly saturated and of uncertain hue. Weintraub's observers adapted to a Ganzfeld which had either a dominant wavelength of 648 mμ, colorimetric purity of 1.00, and luminance of .30 footlambert or a field having the same dominant wavelength and luminance but a colorimetric purity of .50. After 6 minutes exposure to one of these two fields, his observers were exposed to postadapting fields of 648 mμ, 441 mμ, 493 mμ, or 568 mμ, varying in colorimetric purity by steps of .25 between .0 and 1.00, at various luminance levels ranging from .00 to .60 footlambert.

Weintraub stated that with an adapting red of 100% purity, 8 out of 10 observers adapted completely, adaptation times ranging from 2 seconds to 12 minutes. Seven of these observers reported a dark field; the others reported brightness but no hue. With a red-adapting field of 50% purity, 6 out of 10 observers adapted chromatically, times ranging from 2 minutes to 40 minutes. All observers who adapted to the red field of 50% purity also adapted to the red field of 100% purity, and adaptation to the stimulus of lower purity always required more time.

The color experiences initiated by the successive contrast of the postadapting field Weintraub reported to be a function of colorimetric purity but not of luminance: "A change in luminance from adapting field to postadapting field affected perceived brightness, but not perceived hue [p. 559]." Postadapting fields of colorimetric purity equal to or greater than the purity of the adapting field were reported as red; postadapting fields of colorimetric purity less than that of the adapting field were perceived as a complementary blue-green hue.

In view of previous evidence clearly demonstrating the applicability of the principle of color conversion with its emphasis on the luminance dimension as a determinant of perceived hue, Weintraub's results call for additional evaluation. Note, for example, Helson (1938) and Judd's (1960) handling of the Land (1959) phenomenon by color conversion. The present writer is of the opinion that a closer examination of Weintraub's training procedure may change the interpretation of his results. Weintraub exposed observers to a red (648 mμ) adaptation field followed by an achromatic postadaptation field and trained these observers to describe their experience in terms of hue, saturation, and brightness, for example, "hue, red; saturation, 50%; brightness, 10." The investigator did this to teach his observers the judging procedure and to establish a stable color reference system. The color reference system thus established suggests itself as at least a partial explanation of the divergence of Weintraub's results from previous ones. After exposure to one of two red adaptation fields, Weintraub introduced an achromatic field and trained his observers to call the color thus produced by successive contrast "blue-green," an afterimage complement of the red adapting field. Helson (see Footnote 2) suggested as one possible reason for the conflict between his and Weintraub's results the fact that the differences in luminance between the latter's adaptation and postadaptation fields were not great enough to yield afterimage colors. For Weintraub's observers the change was from .30 footlambert to .20 footlambert, whereas the nonselective patches of Helson's 1938 study had reflectance values ranging from .80 to .03. Thus, for Helson's observers the light reaching the eye varied by a ratio of approximately 27–1, whereas the ratio of change for Weintraub's observers was 1.5–1. Also, Weintraub's results might have been different had the shift from adaptation to postadaptation fields been gradual rather than abrupt. The effect of rate of change in illumination is best noted in the dimming phenomenon in which the complementary color response to a chromatic spot on a nonselective background is facilitated by slow, rather than rapid, dimming of the illumination of the spot. Illuminating a nonselective ground with chromatic illuminants, Higbee (1947) found that apertures centrally located in the background showed conversion toward the afterimage complementary of the surround as the ratio of aperture luminance to surround luminance decreased. These results, coupled with those of Helson and Self (1961; cited by Helson, 1964) showing the complementarity of colored shadows to be a function of the ratio of the luminances of the two light sources, support the principle of color conversion and present an explanatory challenge which Weintraub's work fails to meet. At present there remains the challenging research opportunity of demonstrating that Weintraub's results, rather than denying the operation of the principle of color conversion, indicate its operation when the purity of the Ganzfeld is varied.

ORIENTATION IN THE GANZFELD

Upon exposure to the Ganzfeld, observers search hurriedly for something on which to focus the eyes in an effort to orient themselves to the visual field (Miller & Hall, 1962). These investigators found that observers, finding no fixation objects, become aware of vitreous opacities and retinal blood vessels, report gradations of light and shade and, in extreme cases, hallucinations. In addition to these phenomena, the observer soon becomes conscious of an uncertainty as to *where* he is looking. This uncertainty becomes a serious problem if the observer's task is to locate a single object. The observer not only does not know where he is looking, he also does not know where he has been looking previously. Miller and Ludvigh (1960) have shown that it may take as long as 20 seconds to locate an object as large as 6 times the threshold size when no external cues are available. Ludvigh (1936; cited by Miller & Hall, 1962) obtained evidence that with a small section of the cornea anesthetized the eye can be moved mechanically without arousing any sensation of eye movement. He also found (1952; cited by Miller & Hall, 1962) that using false cues can cause an observer to feel that his eyes are in violent motion when they are in fact stationary. From such evidence Ludvigh attributes disorientation in the Ganzfeld to the lack of proprioceptive feedback from the extraocular muscles.

Miller and Hall prefer to interpret orientation in the Ganzfeld as due to the fact that the individual orients himself in relation to some sort of internal bodily coordinate system. They define a displacement threshold as the ability of the individual to place a single test object in the center of the visual field in the absence of all visible detail other than facial features. These investigators produced homogeneous stimulation by means of a clear Plexiglas cylindrical annulus containing a fogging solution through which the observer viewed a uniformly illuminated white field. A test object subtending a visual angle of 25 minutes at the nodal point of the eye was moved in a horizontal arc approximately 11 feet from the observer. This test object was randomly positioned across a 34-degree horizontal path, and observers were instructed to return it to the center of the field. Subjective center was found to be to the left of true center for half the observers and to the right for the other half, average displacements being 2.37 and 2.94 degrees, respectively. Seventy-five percent of the time these observers moved the test object in the appropriate central direction when it was displaced as little as 2.7 degrees from true center. Miller and Hall also provided reference positions at various points in the field to which observers returned test objects. The appropriate reference was presented before each displacement of the test object, and under these conditions the displacement thresholds for different positions averaged 2.0 degrees. These investigators concluded that although the eye may lack an extremely accurate proprioceptive mechanism, the individual appears to possess surprisingly accurate spatial orientation in the absence of retinal cues. Subjective reports of methods used by their observers in locating the center of the field indicate that they resorted to orientation cues from various parts of the body, for example, neck, shoulders, trunk, and hips.

FORM PERCEPTION IN THE GANZFELD

The most primitive level of form perception in the Ganzfeld is that reported by Cohen (1957) who found that the introduction of a brightness gradient into the Ganzfeld at first reversed the usual figure-ground relationship, putting the ground before the figure. Increases in the steepness of the gradient reversed this relationship, producing the usual relation of figure to ground. Cohen also found the spot was judged about 1½ times as large as its real size.

Prolonged homogeneous stimulation seems to result in the failure of the perceptual mechanism to produce a phenomenal field. Cohen

(1960) terms this phenomenon "blank out" and says it is best described as a temporary cessation of vision. This investigator tested observers' ability to discriminate forms during blank out by tachistoscopically presenting (for 1/25 second) figures varying in complexity such as a circle and a star. Time of presentation of the forms coincided with the observer's indication that he was experiencing blank out. With these observers the less complex figures were usually correctly identified during periods of no blank out, but in no case was the observer able to recognize any of the forms during blank out. This lack of ability to recognize the objects persisted even with repeated presentations of the same figure during blank out. At best, observers could describe only a "flash." However, the presence of the forms did, in many instances, seem to terminate the blank out. In the most extreme cases of blank out such a brief presentation of the figures did not produce even the report of a flash. Presentation of the figures for as long as 2 seconds did, however, terminate even the extreme cases of blank out.

Extending his investigation of the effects of homogeneous stimulation, Cohen (1960) presented a series of 16 figures to observers immediately upon their entry into the homogeneous field, after a 90-second adaptation to the homogeneous field, or in a field with a homogeneous center portion and a structured periphery. He found that even a 90-second exposure to a homogeneous field interferes with form recognition. The extent of this interference is somewhat ameliorated by presentation of the series of figures; accuracy of recognition for the first 2 figures is considerably lower than for the remaining 14. This facilitative effect of series presentation is maintained even if exposure of the figures is as short as 1/100 of a second, unless the interval between presentations is extended from 10 to 30 seconds. The most easily recognized figures were the circle, square, diamond, and equilateral and right triangles; most difficult to recognize were the star, pentagon, hexagon, octagon, cross, and the letter X. Rectangles, parallelograms, the letter T, the letter L, and the inverted letter T were of moderate difficulty. Cohen noted that the detrimental effects of the uniform field are most pronounced in the recognition of the more difficult figures; although prolonged exposure to the uniform field reduced the accuracy of recognition for all figures, the greatest loss in accuracy occurred with the more complex figures.

Miller and Ludvigh (1960) investigated the effects of homogeneous stimulation on the acquisition of both stationary and moving test objects. Their Ganzfeld was produced by the cylindrical annulus described previously, and their 5 test objects subtended visual angles of 10.00, 12.34, 17.45, 29.55, and 59.13 minutes of arc at the nodal point of the eye. Stationary targets could be positioned at any one of 29 positions within the field. Moving targets were moved horizontally from one or another of these 29 positions at a rate of .14 degrees per second, a rate previously determined to be below threshold for moving objects. They found that the same general relationship held for all their experimental conditions; length of time required to locate a test object was inversely related to the size of the test object. Acquisition time increased over a range of approximately 3 to 120 seconds, with the increase becoming quite rapid when test-object size was less than approximately 20 minutes of arc. This result was unanticipated in view of the fact that visual acuity previously measured through the fogging solution of the Ganzfeld was found to average 3 minutes of arc for the 3 observers in this study. These authors pointed out the observer's lack of awareness of where he is presently searching or where he has previously searched in the Ganzfeld and noted that this introduces an element of chance into visual search in the Ganzfeld. Calculating probability of acquisition as a function of target size, these authors indicated that the probability of target acquisition does not reach a range of approximately .82 to .92 until target size has increased to approximately 60 minutes of arc. Using as a base the previously determined average visual acuity of 3 degrees, Miller and Ludvigh found it necessary to increase the angular subtense of the target by a factor of 10 over the measured visual acuity in order for acquisition to occur within a relatively short time (approximately 3–5 seconds).

MOVEMENT PERCEPTION IN THE GANZFELD

Cohen (1958a) reasoned that if autokinetic movement under conditions of darkness results from an inadequate visual framework, then similar movement of a figure in an illuminated uniform visual field should occur. Using the adjoining spheres described previously, he exposed observers monocularly to a field containing an 8-centimeter-diameter spot. His subjects observed a series of situations in which differences between the figure and field were systematically varied and, in about 40% of these situations, reported spontaneous autokinetic movement of the spot. The predominance of movement to the right is probably attributable to the fact that Cohen's observers viewed the field monocularly with the right eye. The frequency of reports of movement was not wholly dependent upon the kind of difference between figure and ground, that is, either chromatic differences or intensity differences produced movement. When intensity was gradually increased the previously noted phenomenon of receding ground and approaching figure occurred.

In a subsequent publication, Cohen (1960) reported blurring of contours, transformations of shape, and disappearance of stationary figures in the Ganzfeld. Observers were also found to be relatively insensitive to the movement of objects slowly rotated in the uniform field. When objects were moved at near threshold speeds, movement was often perceived in the direction opposite to the real movement, an extremely confusing effect for the observer when he found that the phenomenal displacement of the object was in a direction opposite to that of his experience of it.

Miller and Ludvigh (1961) noted Aubert's 1886 estimate that the absolute threshold for movement increased by about a factor of 10 when the subject was without a satisfactory frame of reference and Brown's experimental demonstration (1931; cited by Miller & Ludvigh, 1961) that as the background became less and less structured, the phenomenal velocity of a test object decreased. Miller and Ludvigh also reported an experiment in which observers were placed in a Ganzfeld with eyes closed, given a signal to open their eyes, and asked to indicate the cessation of movement of test objects within the field. Two targets, subtending 19.1 and 51.7 minutes of arc at the nodal point of the eye, were exposed for either 5, 10, or 15 seconds and were moved both to the right and left at angular velocities of .30, .56, .71, 1.08, 1.56, and 2.96 degrees per second. "Time delay," the elapsed time between actual stoppage of the target and the observer's report that it had stopped, was the measure of performance. Time delay approached about .5 second for all three durations of movement, for all higher angular velocities. This .5-second value is probably a limiting reaction time since less than .5% of the responses occurred with shorter time delays. As stimulus velocity decreased time delay increased—to a maximum of approximately 6.5 seconds for 5-second exposures, 10.0 seconds for 10-second exposures, and 15.5 seconds for 15-second exposures. No statistical difference was found between delay times as a function of direction of movement. Reported in terms of target size rather than exposure time, the data showed that the larger target produced shorter time delays than the smaller one at angular velocities below 1.50 degrees per second. These authors discussed their results in terms of a hypothesis of "contrast of velocity" whereby the sudden cessation of high apparent speed is held to be more obvious than the sudden cessation of lower apparent speeds. They found, by a chi-square test, that the following formula described their data: $t = (r + 1/av^2)$, where t is time delay in seconds, r is reaction time in seconds, v is the velocity of the test object in angular degrees per second around the nodal point of the eye, and a is the parameter of this inverse proportionality. The adequacy of this formula in describing their data lends mathematical support to Miller and Ludvigh's "contrast of velocity" hypothesis.

ACCOMMODATION IN THE GANZFELD

An empty visual field produces a condition of increased refractive state of the eye similar to the phenomenon known as night myopia. The presence of an empty field also produces fluctuations in the refractive state of the eye. Westheimer and his fellow workers (Campbell & Westheimer, 1959; Westheimer, 1957, 1960; Westheimer, Campbell, & Robson, 1958) have reported a series of studies of fluctuating levels of accommodation produced by empty visual fields. Westheimer (1957, 1960) flashed a beam of .05-second duration into the subject's eye, the beam passing through two fixed positions in the subject's pupil every 10 seconds over a period of 30 to 40 minutes and giving information concerning the refractive state of the eye. The subject's field was either a dark empty field, a bright empty field, or a red fixation spot at optical infinity. The observer was asked to indicate, by pressing the appropriate button or combination of buttons, which of four response categories applied to the presentation of two vertically aligned or misaligned measuring light bars: over accommodation, under accommodation, correct accommodation, or unable to say. These stimulus conditions brought into play positive accommodation usually of the order of .75–1.75 diopters. The accommodative level did not remain fixed throughout an experimental run. Most runs showed no long-term trends, but there were prominent cycle lengths of 1.25 and 2 minutes, and peak to trough amplitudes sometimes ranged up to 1 diopter. Harmonic analysis showed the accommodation time curves had strong components of some frequency bands, but no frequencies characteristic of an observer or a stimulus condition. From these results, Westheimer (1960) concluded that since the refractive state of the eye is unsteady in the presence of a structureless field, it does not seem possible to correct this myopia with a single lens. Flyers have noted that high-altitude flight presents empty field conditions ("sky" myopia) and, consequently, an unstable refractive state. For such conditions Westheimer (1960) noted the probability that an observer, even when wearing the best "average" correction for the task of target detection, may be out of focus by perhaps .5 diopter or more. Brown (1957), working from the hypothesis that optical devices that would produce focus at optical infinity could perhaps correct the unstable refractive state, employed collimated reticles and obtained negative results. He suggested selection of normally hyperopic persons for high altitude target-detection tasks.

According to Koomen, Scolnik, and Tousey (1951), the phenomenon of night myopia was noticed as early as 1883 by Lord Rayleigh. This is nearsightedness developed by the eye in dim illumination, necessitating negative lenses to counteract the decrement in visual acuity. Chin and Horn (1956) mention several bases for this phenomenon: accommodation, increased positive spherical aberration due to pupil dilation in dim illumination, the combination of chromatic aberration and the Purkinje shift, and the fact that a point is not imaged on the retina as a point but forms a caustic along the visual axis near the retina due to spherical aberration.

The magnitude of night myopia is a matter of some dispute; values vary from nearly zero to more than 2 diopters. For example, Otero and Duran (1942), according to Koomen, Scolnik, and Tousey (1951), reported values of 1.5 to 2.0 diopters, while Wald and Griffin (1947) found much more variation from subject to subject. These investigators found for their subjects changes in the power of the eye ranging from + 1.4 to − 3.4 diopters with an average of .59 diopter. Both Koomen, Scolnik, and Tousey (1951) and Helson (1952) drew attention to the fact that such conflicting results are more than likely because various experiments have been carried out at widely differing levels of illumination using a variety of measures.

Paralleling the lack of agreement as to the magnitude of night myopia is the disagreement as to its cause or causes. This disagreement has been largely between Koomen, Scolnik, and Tousey on the one hand, and Otero on the other. The former based their conclusion largely on the fact that night myopia persisted after accommodation was prevented by optical methods and by the use of homatropine, and that it could be produced by means of a lens system having the spherical aberration properties of the eye. Otero (1951) said of his work that ". . . experimental evidence is presented proving that the main cause of night myopia is closely related to the mechanism of accommodation [p. 942]." This experimenter, using a light flash too brief to change the accommodative state of a completely dark-adapted eye, photographed Purkinje images to obtain a measure of the state of accommodation of the eye corresponding to its state of rest. For 3 observers this state proved to be 1.12, 1.22, and 1.24 diopters. Otero maintained that spherical aberration and chromatic aberration contribute .3 and .4 diopter, respectively, to the 1.5 to 2.0 diopters of night myopia he observed. Adding the contributions of these 3 factors to night myopia gives totals approximately equal to the observed night myopia, and Otero concluded that accommodation is its main cause. More recently Chin and Horn (1956) measured the refractive state of the eye in dim illumination and in darkness using an infrared skiascope, and they concluded that the contribution of accommodation to night myopia was generally small.

Helson (1952) pointed out that the Koomen, Scolnik, and Tousey conclusion, emphasizing the role of spherical aberration, is difficult to accept in view of the fact that these authors failed in at least 3 instances to produce spherical aberration in the eyes of subjects. The conflicting results from Otero and Chin and Horn likewise leave uncertain the role of accommodation. Helson's conclusion in 1952 that the problem was not closed is equally applicable at this time.

GANZFELD "BLANK OUT" AND ITS ELECTROPHYSIOLOGICAL CORRELATES

In a series of 3 studies of vision in the Ganzfeld, Cohen (1956, 1957, 1960) reported for some observers a temporary cessation of visual experience; this phenomenon he termed "blank out." It was found in 5 out of 16 observers, 2 of whom reported it almost every time they were exposed to the uniform field. The experience was not simply the presence of a dark, undifferentiated field, but rather a "complete disappearance of the sense of vision for short periods of time." Cohen (1960) found various degrees of blanking out. For example, the observer may experience a feeling of light with no external reference. Or, in the most extreme cases, he may experience a "complete absence of seeing," this latter extreme generally occurring only after exposures of 10–20 minutes. Observers reported uncertainty as to whether or not their eyes were open; they were also unable to control their eye movements. Although blinking and eye movement usually brought about the return of the visual field, this did not occur with the most extensive losses of vision.

Cohen (1960) suggested that persistent uniform stimulation without a structured field results in the failure of the perceptual mechanism to produce a phenomenal field, a possibility supported by the fact that reports of blank out were found to be associated with the return of bursts of alpha activity. Cohen pointed out that alpha patterns in the EEG record are usually associated with the absence of visual stimulation. The return of alpha activity during blank out, he suggested, would indicate a functional similarity between no stimulation and prolonged uniform stimulation. Individual differences in susceptibility to blank out Cohen found to be correlated with individual EEG records; high alpha activity in the absence of visual stimulation was associated with more susceptibility to blank out. Cohen and Cadwallader (1958) found supporting data for this conclusion.

Cohen's suggestion generated further research by Tepas (1962), who exposed 3 subjects with normal vision and normal EEG's to a white, uniformly illuminated sphere for 5-minute periods followed by 3-minute rest periods. Electrical activity was recorded from six standard EEG electrodes. Subjects were initially instructed to describe what they saw as accurately as possible. Once the subject verbalized experiencing blank out, he was instructed to close a microswitch which produced a signal on the EEG record. On all experimental sessions, alpha scores were obtained at the beginning and end of the session under "eyes closed" and "eyes open" conditions. Tepas found that all subjects experienced blank out and that the alpha present during Ganzfeld viewing was intermediate between "eyes closed" and "eyes open" conditions for all observers. There was a significant absence of saccadic eye movement or blinking just prior to onset of blank out and during the experience, but the end of the experience coincided with eye movements or blinking to a statistically significant degree. Also, an increase in the number of viewing trials increased the amount of alpha present but did not alter the blank-out/non-blank-out alpha difference.

Data from a typical high alpha subject indicated that, irrespective of the general level of ongoing alpha activity in the brain, there was significantly more alpha during blank out than when blank out was not reported. There was also a general increase in the level of alpha activity as the experimental session progressed; however, the amount of blank out reported had become relatively stable at 10 to 15% by the middle experimental session and did not increase thereafter with the level of alpha activity. This finding indicates that amount of blank-out time need not be merely a function of the general level of alpha activity. Tepas' data showed a negative relationship between amount of eye-movement potentials and amount of blank out reported, causing him to conclude that the hypothesis that high alpha activity is an important indicant of blank-out susceptibility was not strongly supported.

More detailed specification of the factors contributing to the relationship between blank out and alpha activity deserves the continued interest of physiological psychologists, as do the relationships between the factors contributing to night myopia. These two problems, along with the possibility of demonstrating the operation of the principle of color conversion when colorimetric purity is varied, are likely to be rewarding ones for interested researchers.

REFERENCES

BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology,* 1954, **8,** 70–76.

BROWN, J. F. The visual perception of velocity. *Psychologische Forschung,* 1931, **14,** 199–232.

BROWN, R. H. "Empty-field" myopia and visibility of distant objects at high altitudes. *American Journal of Psychology,* 1957, **70,** 376–385.

CAMPBELL, F. W., & WESTHEIMER, G. Factors influencing accommodation responses of the human eye. *Journal of the Optical Society of America,* 1959, **49,** 568–571.

CHIN, N. B., & HORN, R. E. Infrared skiascopic measurements of refractive changes in dim illumination and in darkness. *Journal of the Optical Society of America,* 1956, **46,** 60–66.

COHEN, W. Comparisons of homogeneous Ganzfelds with Ganzfelds containing simple figures. *Dissertation Abstracts,* 1956, **16,** 1510–1511.

COHEN, W. Spatial and textural characteristics of the Ganzfeld. *American Journal of Psychology,* 1957, **70,** 403–410.

COHEN, W. Apparent movement of simple figures in the Ganzfeld. *Perceptual and Motor Skills,* 1958, **8,** 32. (a)

COHEN, W. Color perception in the chromatic Ganzfeld. *American Journal of Psychology,* 1958, **71,** 390–394. (b)

COHEN, W. Form recognition, spatial orientation, perception of movement in the uniform visual field. In A. Marris & E. P. Horne (Eds.), *Visual search techniques.* (Pub. 712) Washington: National Academy of Science-National Research Council, 1960. Pp. 119–123.

COHEN, W., & CADWALLADER, T. C. Cessation of visual experience under prolonged uniform visual stimulation. *American Psychologist,* 1958, **13,** 410. (Abstract)

GIBSON, J. J., & WADDELL, D. Homogeneous retinal stimulation and visual perception. *American Journal of Psychology,* 1952, **62,** 263–270.

HELSON, H. Fundamental problems in color vision: I. The principle governing changes in hue, saturation, and lightness of nonselective samples in chromatic illumination. *Journal of Experimental Psychology,* 1938, **23,** 439–476.

HELSON, H. Vision. *Annual Review of Psychology,* 1952, **3,** 55–84.

HELSON, H. *Adaptation level theory.* New York: Harper & Row, 1964.

HELSON, H., & JEFFERS, V. B. Fundamental problems in color vision: II. Hue, lightness, and saturation of selective samples in chromatic illumination. *Journal of Experimental Psychology,* 1940, **26,** 1–27.

HELSON, H., & JUDD, D. B. A study of photopic adaptation. *Journal of Experimental Psychology,* 1932, **15,** 380–398.

HELSON, H., & SELF, H. C. A study of coloured shadows. Paper read at Maxwell Color Centenary, London, 1961.

HIGBEE, R. V. Hue and saturation of aperture colors as a function of the composition and luminance of the surrounding field. Unpublished doctoral dissertation, Bryn Mawr College, 1947.

HOCHBERG, J. E., TRIEBEL, W., & SEAMAN, G. Color adaptation under conditions of homogeneous visual stimulation (Ganzfeld). *Journal of Experimental Psychology,* 1951, **41,** 153–159.

ITTELSON, W. H., & AMES, J. A. Accommodation, convergence, and apparent distance. *Journal of Psychology,* 1950, **50,** 43–62.

JUDD, D. B. Appraisal of Land's work on two-primary color projections. *Journal of the Optical Society of America,* 1960, **50,** 254–268.

Koffka, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

Koomen, M., Scolnik, R., & Tousey, R. A study of night myopia. *Journal of the Optical Society of America,* 1951, **41**, 80–90.

Land, E. H. Experiments in color vision. *Scientific American,* 1959, **200**, 84–94, 96, 99.

Ludvigh, E. J. Is ocular proprioceptive sense concerned in vision? *Archives of Ophthalmology,* 1936, **15**, 1037–1049.

Ludvigh, E. J. Possible role of proprioception in the extraocular muscles. *Archives of Ophthalmology,* 1952, **48**, 436–441.

Metzger, W. Optische Untersuchungen am Ganzfeld: II. Zur phanomenologie des homogenen Ganzfelds. *Psychologische Forschung,* 1930, **13**, 6–29.

Miller, J. W., & Hall, R. J. The problem of motion perception and orientation in the Ganzfeld. In M. A. Whitcomb (Ed.), *Visual problems of the armed forces.* Washington: National Academy of Science-National Research Council, 1962. Pp. 14–20.

Miller, J. W., & Ludvigh, E. Time required for detection of stationary and moving objects as a function of size in homogeneous and partially structured visual fields. In A. Marris & E. P. Horne (Eds.), *Visual search techniques.* (Pub. 712) Washington: National Academy of Science-National Research Council, 1960. Pp. 170–180.

Miller, J. W., & Ludvigh, E. The perception of movement persistence in the Ganzfeld. *Journal of the Optical Society of America,* 1961, **51**, 57–60.

Otero, J. M. Influence of the state of accommodation on the visual performance of the human eye. *Journal of the Optical Society of America,* 1951, **41**, 942–948.

Otero, J. M., & Duran, A. Continuación del estudio de la miopia nocturna. Notas II y III. *Anales de física y química,* 1942, **38**, 236–244.

ten Doesschate, G. Vision in an empty field. *Psychological Abstracts,* 1961, **35**, 565.

Tepas, D. L. The electrophysiological correlates of vision in a uniform field. In M. A. Whitcomb (Ed.), *Visual problems of the armed forces.* Washington: National Academy of Science-National Research Council, 1962. Pp. 21–25.

Wald, G., & Griffin, D. R. The change in refractive power of the human eye in dim and bright light. *Journal of the Optical Society of America,* 1947, **37**, 321–336.

Weintraub, D. J. Successive contrast involving luminance and purity alterations of the Ganzfeld. *Journal of Experimental Psychology,* 1964, **68**, 555–562.

Westheimer, G. Accommodation measurements in empty visual fields. *Journal of the Optical Society of America,* 1957, **47**, 714–718.

Westheimer, G. Accommodation levels in empty visual fields. In E. P. Horne & M. A. Whitcomb (Eds.), *Vision research reports.* (Pub. 835) Washington: National Academy of Science-National Research Council, 1960. Pp. 21–23.

Westheimer, G., Campbell, F. W., & Robson, J. B. Significance of fluctuations in accommodation. *Journal of the Optical Society of America,* 1958, **48**, 669.

(Received April 5, 1965)

Psychological Bulletin
1965, Vol. 64, No. 4, 259–309

GROUP COHESIVENESS AS INTERPERSONAL ATTRACTION:
A REVIEW OF RELATIONSHIPS WITH ANTECEDENT AND CONSEQUENT VARIABLES[1]

ALBERT J. LOTT
University of Kentucky

AND

BERNICE E. LOTT
Kentucky State College

The cohesiveness of small groups is defined in terms of intermember attraction and the rationale for such an approach is discussed. The empirical literature, restricted primarily to investigations published between 1950 and 1962, is reviewed with the aim of evaluating the status of variables hypothesized as having antecedent or consequent relationships with interpersonal attraction. To this end, studies from diverse fields, for example, group dynamics, personality, and learning, are brought together and categorized. Theoretical positions concerned with the development of liking between persons and the effects which liking has upon subsequent behavior are also examined by specifying predictions from systematic formulations and comparing them with the research data. The major intent of this paper is to document relationships which have been clearly established and to identify those which are still equivocal or unexplored.

The authors are concerned here with research and theory relevant to the concept of cohesiveness. Their starting point in reviewing and categorizing the literature is the definition of cohesiveness as "that group property which is inferred from the number and strength of mutual positive attitudes among the members of a group [Lott, 1961]." While it may be said that this definition merely focuses on one of several components of cohesiveness, there is good reason to assume that interpersonal attraction, liking, or positive attitudes among group members, is central to the cohesiveness of small groups, whether other factors are also viewed as independently relevant or not.

Most researchers in the area have subscribed to the position introduced by Festinger, Schachter, and Back (1950) that the cohesiveness of a group, that is, the desire of individuals to maintain their membership in a group, is contributed to by a number of independent forces, but most investigations have focused on one force, intermember attraction. When investigators have desired to manipulate the cohesiveness of groups, regardless of their nominal definition of the concept, the operations performed have typically involved telling the members of some groups that they would probably like each other, be congenial, etc., while telling others just the opposite. Further, in measuring the cohesiveness of experimental or naturally existing groups, some sociometric device is generally utilized to determine how much each member likes or is attracted to the other members. For the most part the other forces which are assumed to determine, or be reflected by, the cohesiveness of a group have received less attention, although some investigators have manipulated variables other than interpersonal attraction in varying group cohesiveness, and some have measured it along different dimensions. Defining cohesiveness in terms of interpersonal liking is, therefore, only relatively arbitrary.

Tagiuri (1958) has noted that

> Factor analytic studies of mutual ratings by members of small groups concur on the presence of three basic factors: influence and initiative, task competence, and like-dislike. While the first two are not applicable to every group, the last one is always present. . . . [and] other things being equal . . . the category of like and dislike "packages" most of the determinants of interaction [p. 317].

In a more general vein, we may cite Bonner (1959):

---

[1] This paper was prepared under a grant from the Cooperative Research Program of the Office of Education, United States Department of Health, Education, and Welfare. We are grateful to Theodore M. Newcomb for his critical reading of an earlier version of the manuscript.

> If we analyze group cohesiveness . . . in terms of a group's attractiveness for its members, we are confronted by the obvious fact that without at least a minimal attraction of members to each other a group cannot exist at all [p. 66].

In the present discussion, we shall be concerned only with studies of interpersonal attraction among group members and the relation of interpersonal attraction to other group process variables. How do persons in coacting or interacting groups get to like one another? And, once they do to some measurable degree, what are the consequences of this state of affairs for the group or for the behavior of the members? We want to take stock of the answers which existing research has already given to these questions and then compare the empirical findings with theoretical statements, that is, with predictions which have been, or can be, made from some of the more systematic formulations of group behavior.

Considered in this review are investigations of real or simulated interaction among persons who are in association with one another on a relatively voluntary basis. Omitted are studies of special populations like prisoners or psychotics. Another limitation is provided by date of publication; our interest is in the recent research literature and the majority of the investigations considered here have been reported between 1950 and 1962. Other reviews have provided adequate coverage of the earlier literature (e.g., Cartwright & Zander, 1953, 1960; Kelley & Thibaut, 1954; Lindzey & Borgatta, 1954; Riecken & Homans, 1954; Roseborough, 1953; van Bergen & Koekebakker, 1959). We have also, for the most part, omitted unpublished studies and technical reports. Inevitably we will have bypassed investigations which some consider of utmost relevance despite our attempt at thoroughness within the limitations mentioned above.

The number of relevant and near-relevant papers is staggering, but while a small number of phenomena or relationships have been demonstrated over and over again, other relationships remain as hints, unexplored, or equivocal. We can hopefully document both the former and latter, helping researchers to move away from the known in the direction of either more refined investigations of clear relationships or concentration on those which have not been certainly established. We will not deal with problems of methodology and measurement, but will focus on what has been shown to hold over a variety of situations, using various techniques to manipulate conditions and innumerable devices to infer changes in dependent variables.

ANTECEDENTS OF LIKING: EMPIRICAL RELATIONSHIPS

The research considered here will be discussed under classes of variables which have been empirically linked to the consequent attraction of one individual, A, to another individual, B. Within each category may be found a number of different types of investigation which can be roughly classified as follows:

1. Experimental manipulation of Variable *X* (or Variables *X*, *Y*, etc.) with consequent direct measurement of the dependent variable, liking, or some other dependent variable which is assumed to reflect interpersonal liking, for example, attraction to the group. In some cases this assumption will have been made by the original investigator; in others, by the present writers.
2. Natural groups which are known to differ in the degree of interpersonal liking existing among the members are compared on other variables to determine relationships with liking. In some cases it is reasonable to assume that these other variables antedate interpersonal feelings within the groups studied while in other cases the question of "which came first" cannot be satisfactorily answered.

We are primarily concerned with liking as measured by direct choice of A for B, or as measured by A's rating of B on a scale, where the criterion for choice or rating is of a general social nature as opposed to one of a specific work or task-oriented nature. Studies concerned with the latter have been considered if they were viewed as having relevance to more general attraction.

Interaction-Propinquity

A sizable number of studies have supported the general hypothesis that interpersonal attraction is a positive function of interaction.

Bovard (1951a, 1956a, 1956b) has shown that positive affect toward individual group members and toward the group as a whole will increase significantly more in group-centered classes where verbal interaction is encouraged than in leader-centered classes, and that the relationship between average interaction within groups and average affect ratings of individual members is approximately linear. McKeachie (1954), too, found that students in group-centered classes differed reliably from students in leader-centered classes in the direction of greater liking for their group. Propinquity in boarding school classes was reported to be very reliably linked with liking choices by Maissonneuve, Palmade, and Fourment (1952). Similarly, in college classes, seat neighbors were found to become acquainted with one another significantly more often than non-neighbors (Byrne & Buehler, 1955), and the intensity of the relationship formed (i.e., whether best friend, friend, or acquaintance) has been reported to be greater among the former than the latter (Byrne, 1961a).

Studies of other natural or semi-natural settings, outside of the college classroom, have also yielded evidence in support of the relationship between opportunity for interaction and attraction. Gullahorn (1952) observed and interviewed clerical workers in an office of a large corporation. He found that distance was the single most important factor determining rate of interaction between any two employees, and that there was a tendency for persons who interacted frequently to develop "sentiments of friendship." In a housing project for married veteran university students, the two major factors affecting friendship were found to be sheer distance between houses and the direction in which the houses faced (Festinger, 1953). In another study, sorority girls living together chose best friends among themselves (house residents) significantly more often than they chose sorority sisters who lived outside of the "house" in town (Willerman & Swanson, 1952). The sociometric preferences of bomber crew members were shown to be reliably predictable from a knowledge of their location and function on the plane, that is, from a knowledge of who they had the greatest opportunities to interact with (Kipnis, 1957). In a summer camp, Sherif and Sherif (1953) divided boys into two groups by breaking up budding friendships. As boys in each of the new groups worked and played together, friendship choices were reversed or shifted away from the boys preferred earlier toward boys within the new group. A subsequent study (Sherif, White, & Harvey, 1955) duplicated these results. Similarly, Heber and Heber (1957) have reported that elementary school children increase their favorable ratings of each other on a social distance scale following common small-group experience, and Zander and Havelin (1960) found that among three-man groups of initial strangers in the Norwegian Navy, men preferred to remain in their own groups following a period of interaction. Both of the last mentioned investigations also demonstrated that the interaction-attraction relationship is subject to variation in degree as a function of conditions such as success or failure and task competence of group members.

Even in a competitive situation, individuals, following interaction, have been found to become more favorable in their ratings of the traits of both their own teammates and their opponents (Wilson & Miller, 1961), although the increase in favorableness for teammates and opponents is not the same under all conditions. Finally, the reduction of prejudice by white persons toward Negroes has been reported to covary with increased contact between persons of different color in situations as diverse as a meat packing plant (Palmore, 1955), a housing project (Deutsch & Collins, 1958), and a university classroom (J. H. Mann, 1959).

That sheer contact may be a necessary but not sufficient condition for attraction has been shown by other investigators. Festinger (1953), for example, has discussed a housing project in which few group memberships and little social life existed among residents. Here residents, who felt "forced" to live in the project by circumstances beyond their control (housing shortage), were largely negative in their attitude toward their community and neighbors. Gundlach (1956) found, among a sample of white women workers who were members of left-wing unions with strong fair employment practices, persons in one subgroup with high education and high social aims who were in contact with Negroes of similar education and aims to be most hostile and derogatory toward Negroes. This subgroup was also most antiunion. The other white union members, on the other hand, exhibited far less prejudice than that reported for samples of the general population.

In two controlled studies, Stotland, Cottrell, and Laing (1960) and Stotland and Cottrell (1962) found no significant relationship between interaction and desire to get to know one's fellow group members. The interaction in these investigations conveyed no information about members' characteristics, was nonverbal, and only faces or heads of members were visible. This was not the case, however, in another investigation (of sorority girls residing in the same house) where only a slight correlation was obtained between ranks on a liking scale and frequency of interaction (Backman & Secord, 1962).

It seemes likely, from the evidence available, that interaction, to produce attraction, should take place in a relatively neutral atmosphere in which there are opportunities for verbal communication or the observation of one another's behavior.

Special Characteristics of the Group Situation or "Atmosphere"

Cooperation. Stendler, Damrin, and Haines (1951) observed groups of second-grade children working for individual reward and for group reward. Only under the latter condition did positive interactions exceed negative ones. Groups of fourth-grade children were studied by Phillips and D'Amico (1956). In one condition rewards were distributed according to individual contributions while in another rewards were shared equally. Attraction to members of one's group, measured sociometrically, increased in more groups of the latter kind than in those of the former. Further, in those competitive groups where rewards were distributed fairly evenly because members were well matched, the level of intermember attraction either remained as it was or increased. Increased liking, acceptance, or friendliness among subjects who have cooperated on a task, in contrast to those who have competed, has also been reported by Gottheil (1955) working with eighth-grade students, by J. R. P. French (1951) who studied adults attending a laboratory for training in group relations, and by Deutsch (1960) who varied the conditions under which reward was attained within college classroom groups.

In a recent study by Solomon (1960), pairs of subjects played a game while each partner was in a separate cubicle, unable to see the other, and communicated via electric switches. In such a situation, too, subjects who received a cooperative strategy from their partners (i.e., one promoting maximum mutual gain) tended to develop greater liking for them than did subjects who received a noncooperative strategy.

Negative findings have been reported by Rosenbaum (1959) who had pairs of male undergraduates discuss together for 20 minutes their choice of an ideal college. At the end, cooperating pairs (defined in terms of both members being eligible for reward as opposed to just one) differed significantly from competing pairs on a measure of assumed similarity of personality traits, but not on a social distance scale of desired relationship with partner. These findings contradict most of the others reported, but it is not clear what the responsible conditions might be. In a study by Myers (1962) members of competitive rifle teams were found to increase their esteem ratings of their teammates more than members of noncompetitive teams, but in this case the individual members of competitive teams were not working against each other, but were working together against other teams.

In summary, the hypothesis that persons will like one another if they have worked together for common rather than for mutually exclusive or individual ends is generally supported.

Democracy. White and Lippitt (1960) have reported that, in the classic experiment on social climates conducted by Lewin and his co-workers in 1939, when the boys' groups were supervised by a "democratic" leader there was more friendliness among the boys than when they were supervised by an "autocratic" or "laissez-faire" leader. There was, further, greater group mindedness, more mutual praise, and more spontaneous subgroups of five and four—one representing the greatest possible unity in a five-person group. No sociometric data were obtained in either of the two experiments comprising the investigation, and judgments of friendliness are based primarily on the analysis of conversations. It is regrettable, then, that so few other studies have been designed to specifically test the hypothesis that between similar groups, differing only in leadership style, members of democratic groups will like each other more than will members of autocratic ones.

Inconclusive data bearing on this question were obtained by Rehage (1951) in a study of eighth grade social studies classes. In one class planning was done by pupils and the teacher, while in the other it was done by the teacher alone. Students in the two classes had been matched on IQ scores and on an acceptance-rejection index. At the end of a year there was no difference between the two classes in acceptance-rejection indexes of pupils, but there was a difference in general group structure: well-defined subgroups present at the beginning tended to persist in the traditional class, whereas friendship patterns had been "modified perceptibly" in the other.

Berkowitz (1953) has also reported relevant data from an investigation involving observation of seventy-two actual conference groups in government and industry and the testing and interviewing of individual participants. Leader permissiveness in discussion content was found to correlate positively with an index of group cohesiveness (which included a measure of personal liking among members), especially in groups characterized as having "urgent problems" to solve. Further analyses indicated, however, that in groups which were permissive with respect to content the cohesiveness index correlated positively with leader control of group process and with functional differentiation of leader's role, and negatively with percent of participation and problem solution by members. Individuals in the groups studied by Berkowitz shared, apparently, "an expectation . . . that the designated leader, the chairman, is to be the major behavioral leader in the group," and the more this was true, the more satisfied were the members with their conference.

That there is a clear and simple relationship between democratic participation and consequent liking among group members seems doubtful. It is more likely that the degree to which group members' role expectations are fulfilled is the more significant variable in producing satisfaction with the group and with each other.

Acceptance by Others

A number of studies have followed the general procedure of manipulating conditions so that A believes himself to be liked (or not liked) or accepted to some degree by B, and A's attraction toward B is subsequently measured. In an investigation by Backman and Secord (1959), for example, subjects were given the names of others who would most *probably* like them (as judged by the experimenter's evaluation of personality test data). At the first interaction session the number of these "probable likers" chosen as desired "team partners" was significantly above chance. At subsequent interaction sessions (during which the subjects got to know one another) this initial effect disappeared. In variations of this procedure, Kelley and Shapiro (1954) and Dittes and Kelley (1956) found that subjects made aware of positive ratings of themselves presumably made by their fellow group members differed significantly from others made aware of negative ratings in their desire to remain in the group or to work with the other members. The dependent measures used in the earlier study included ratings of individual co-workers. Simiarly, Dittes (1959), who permitted subjects to see fictitious ratings of themselves by others, found that the more accepted members were significantly more attracted to the group qua group and that this was more true of persons with low as opposed to high self-esteem.

In a study of the consequences of rejection, Pepitone and Wilpizeski (1960) demonstrated that in the no rejection condition, where the subject believed himself to be significantly more liked by his group members (two confederates) than in the rejection condition, he also tended to like his fellow members reliably more. Pepinsky, Hemphill, and Shevitz (1958) manipulated rejection and acceptance by having stooges react negatively or positively to attempts at leadership by subjects. Under the first condition, the confederates received a disproportionate share of negative sociometric choices while under the second, favorable choices were randomly distributed between confederates and other subjects, and the groups were significantly higher in "morale." Harvey (1962) has shown a relationship between subjects' reevaluations of friends and strangers and the ratings presumably received from the friends and strangers: "agreement or near agreement of the source's with subject's self-ratings resulted in subject becoming more positive toward the source, particularly the stranger source. . . ." Also concerned with rejection, Snoek (1962) compared its effects when of an invidious and noninvidious nature, the latter being impersonal and not implying negative evaluation of the rejected person. Subjects exposed to the latter condition showed a greater reduction in attraction to the group as measured by the question, "Do you want to remain a member?" but the reason for rejection did not make a difference on a measure of rejected subjects' desires for contact with individual members. The reasons for rejection differed not only on the invidious-noninvidious dimension, but also in terms of a subject's expectation that he might join the group (bridge club). When the rejection was invidious it was based on personal qualities and skill (open to change), but when non-invidious it was based on the subject's sex, that is, the other group members preferred to have a girl. Utilizing a procedure designed to induce not rejection, but simply a "fear of rejection" in some groups and not in others, de Charms (1957) found a tendency (nonsignificant) for subjects in the former to find their groups less attractive, as inferred from responses to a questionnaire.

Some studies have manipulated acceptance by varying status within the group. In a role-playing situation, Zander and Cohen (1955) found that subjects playing high-status persons were significantly more attracted to their groups than those playing low-status persons, and the former also differed from the latter in having perceived themselves as making a good first impression and having influenced the group. Berkowitz and Macauley (1961) reported that subjects who believed themselves to have been elected by other group members as discussion leaders tended to be more highly attracted to their fellow members both as social companions and work partners if they also believed that their high status could drop.

In a questionnaire study of staff members in a child welfare agency, Jackson (1959) obtained some evidence that professional personnel selected by co-workers as being the "most valuable" in terms of their contributions tended to be more attracted to their groups and its members, especially if they were in high contact with one another. Also utilizing a natural environment, Newcomb (1956) studied seventeen men, initially strangers, housed together in a dormitory. Throughout the period of interaction there was a positive relationship between liking of others and believing oneself to be liked by them. Similar findings were reported earlier by J. R. P. French (1951) and Tagiuri (1952). French found that among the bases on which friendship choices were made among group members was the extent of actual and expected reciprocation of choice, and Tagiuri, after investigating about thirty different groups, reported that the "relationship between perception of affect and affective response has been found to obtain generally, irrespective of the size of the group, the sex and the age of the subjects. . . ." Whether A's perception that he is positively evaluated by B antedates his liking of B is not clear from these particular studies, but the controlled experiments which have been cited amply support the conclusion that A will like B if B likes or accepts A. That, as we shall see later, one of the consequences of A liking B seems to be a tendency for A to perceive himself as also being liked by B lends additional significance to the importance of mutuality or reciprocation.

Frustration–Threat

Sherif and Sherif (1953) reported that the consequence of frustrating and competitive intergroup relations for boys at a summer camp was to solidify "in-group belongingness and solidarity . . . and to strengthen in-group friendships." Earlier, Wright (1943) had observed the same general phenomenon. Pairs of children were subjected to frustration by the experimenter with a consequent increase in cooperative behavior and a decrease in time spent in conflict. No direct measure was obtained of friendship or liking. In another early study, the well-known "social climate" investigation by Lewin (cited by White & Lippitt, 1960) and his co-workers, the autocratically led, submissive, groups were found to surpass all others in the proportion of friendly remarks among the boys. Two other studies in which no direct data on intermember liking were reported are also relevant. Lanzetta, Haefner, Langham, and Axelrod (1954) and Lanzetta (1955) found that groups of reserve officer trainees working, in one study, under the threat of "evaluation" and, in the other, under the stress of time and negative comments by the experimenter were more sociable, cooperative, and friendly than groups working on identical tasks but under no threat. More recently, Myers (1962) reported that individuals in three-man teams facing "adversity," that is, competition against other teams, pulled together and rated one another higher than did individuals in noncompetitive teams. Despite the fact that an absent member reduced his team's score on the day of his absence, absent members, too, were rated higher by members of competing groups. Members of competing groups also perceived themselves as being more accepted by their teammates than did members of noncompeting teams.

The findings of some investigators indicate that the shared threat-attraction relationship is influenced by other variables. Hamblin (1958), for example, demonstrated that group integration does not increase under threat or crisis when a likely solution requiring cooperation is unavailable to the group members. Some groups were subjected to arbitrary rule changes while playing a game; others were not. The former groups were observed to differ significantly from the latter in the following areas of behavior, exhibited after the "crisis": less helpfulness among members, less praise of others, and greater antagonism. No data were obtained, unfortunately, on groups subjected to threat where a likely solution *was* available. Data reported by Mann and Mann (1959) may be interpreted as supporting Hamblin's hypothesis. Classroom groups, meeting four times a week for one hour over a three-week period, were organized as task-oriented study groups to discuss assigned lists of reading or as free discussion groups. Ratings of the members' desirability as friends increased in the former groups and decreased in the latter where, according to observers, the subjects were frustrated and angered by the indefiniteness of their situation. As in Hamblin's study, perhaps these subjects saw "no way out."

That the threat shared by individuals should be one emanating from an *external* source in order for increased liking of one another to follow is suggested by the findings of Pepitone and Kleiner (1957). High status groups of boys at a summer camp (i.e., winners of two out of three tournament games) were told either that they would probably lose the tournament (threat) or that they would probably win. Following the threatened loss of status there was no change in the number of positive choices made among teammates while among those told they would probably win there was an increase in positive choices. In this study threat was operationalized by a prediction of failure made by a source external to the group but failure would necessarily result from poor playing on the part of the boys themselves.

It is apparent from the studies cited that definition or manipulation of "threat" varies from one investigation to the next. It has, for example, been equated with intergroup competition, with frustration or negative treatment, and with threat of failure. In other research (e.g., Burnstein & McRae, 1962) poor evaluation of group performance on a task has been used to induce a condition of "shared threat." Such studies, in which actual success or failure constitutes the major independent variable, will be discussed separately in a subsequent section. Despite differences in the operations used to create a stressful situation, the evidence examined provides a fairly consistent pattern. We suggest that attraction among individuals will be found to increase when their common threat stems from an external source (i.e., is not a function of their own lack of skill), when there exists the possibility that cooperative behavior may reduce or eliminate the threat, and when single individuals cannot escape from either the group or the threat.

Status

Status Similarity. Thibaut (1950) manipulated the status of teams of boys by assigning differential roles to each team in a series of games, and the boys subsequently chose teammates for a further game. Those in high-status groups (especially peripheral or less popular boys) showed a significant increase in proportion of own-team choices. Popular or central members of low status groups which had been unsuccessful in an attempt to improve their status also increased their proportion of own-team choices. We shall consider this latter finding again in our discussion of the effects of the success-failure variable on interpersonal attraction. Our concern here is with the consequences of status defined by operations other than success or failure on a task. Further evidence appears in a report by Festinger (1953) in which he compared a housing project for married veteran students (Westgate) with a government project outside of a large city (Regent Hill). The inhabitants of the two projects were described as differing sharply in the evaluation of the prestige and desirability of their respective communities. The majority of those living in Regent Hill resented having to be there (because of the postwar housing shortage) and thought of their neighbors as lower class. The students of Westgate, on the other hand, were living among preprofessional peers and had a generally positive attitude toward their community. It is not surprising, then, that many more informal and formal groups were established and friendships developed among neighbors in Westgate than in Regent Hill. In another field study of industrial work groups, Seashore (1954) found a small but significant positive correlation between perceived status of one's job and an index of group cohesiveness (which did not include a measure of intermember liking). Since perceived status of job was determined by asking individuals whether they had a "good" job, it is not clear whether the obtained findings reflect a relationship between cohesiveness and job prestige or between cohesiveness and job satisfaction.

Other relevant research has been done on groups composed of members differing from one another in status. Individuals high in professional prestige, attending a one day conference, were found to prefer others like themselves for friendly interaction over individuals of lower status (Hurwitz, Zander, & Hymovitch, 1960). In a laboratory situation where status was manipulated by the desirability of job performed, subjects in high-status but mobile positions tended to accept persons at their own level and reject others (Kelley, 1951), although there were no significant differences between conditions in choice of best-liked person. Cohen (1958) has reported similar findings to Kelley's regarding high-status subjects. Jennings (1950) also found that mutual choice was greatest among the overchosen or high-status girls (the leaders) on a "live and work with" criterion question. Mutual rejection was also greatest among these girls, suggesting the operation of competition or rivalry (Homans, 1961) under the same conditions which produce attraction.

The above studies generally support the conclusion that members of the same high-status group or members of mixed-status groups occupying high positions will tend to like one another.

Status Dissimilarity. That the tendency for high-status persons to be attracted to others of high status (considered above) is probably independent of a similarity-attraction relationship is suggested by findings which indicate a general preference for high-status persons. Hurwitz, Zander, and Hymovitch (1960), for example, found that individuals low in professional status liked those of higher status more than they liked each other. Data on direction of communication support the sociometric findings. Masling, Greer, and Gilmore (1955), working with military men, also found status to be highly related to sociometric ratings, and Tagiuri, Kogan, and Long (1958) reported that college preparatory students tended to choose (for a roommate), without perceiving themselves chosen in return, individuals with high choice status, while they perceived themselves as being chosen by individuals of lower status. In the previously mentioned Kelley study (1951), the low-task-status but potentially mobile subjects exhibited friendly behavior to high-status subjects, as judged from an analysis of written messages; this was not true, however, of low-status, nonmobile individuals. Cohen's (1958) replication of this study obtained similar data and, in addition, found that nonmobile, low-status subjects sent significantly more cohesiveness-enhancing comments to their own subgroup than did mobile, low-status subjects. The nonmobile, low-status subjects may have experienced a condition of common frustration which, perhaps, was a more powerful variable than status in influencing the development of positive affect toward other group members. Findings from another study (Thibaut & Riecken, 1955) suggest that some people may be more prone than others to be influenced by relative status in their acceptance or rejection of persons. Men high in authoritarianism, after experiencing aggression at the hands of an instigator, rejected him less if his status was high and rejected him more if his status was low than did less authoritarian men.

A number of investigations have found that individuals tend to like persons who have influenced them, and it seems reasonable to include such findings in the present discussion since the direction of influence is most usually from a position of high to a position of lower status. Whether influence precedes attraction, or vice versa, is, however, not clear, since the data have come from studies in which influence has not been manipulated. Perlmutter (1954), for example, has reported that within adult discussion groups the greater the perceived influence of a member, the more desirable were the traits assigned to him; in some groups a significant positive correlation was also obtained between the valence of a group member with respect to "most prefer working with" and his perceived influence. Earlier, Lippitt, Polansky, and Rosen (1952), in a replication of an earlier study, found that boys in a summer camp situation tended to like those boys to whom they also attributed high power (or influence potential). Tagiuri and Kogan (1960) have reported positive correlations between liking of others and feeling influenced by them as well as between liking of others and feeling influential over them. The latter finding, together with evidence from Kelley (1951), suggests another phenomenon toward which additional research might well be directed. Kelley found that high-status, nonmobile subjects, that is, those not threatened by loss of status, exhibited as much interlevel friendliness as did subjects in the nonhierarchic control condition.

Status Congruence. Adams (1953) found that in bomber crews where members were in the same status relationship to one another on a number of different dimensions, for example, age, education, role prestige (where, in other words, status congruency was high), there were more friendships than in crews where status congruency was low. Similarly, in a study of supermarkets by Clark (cited by Homans, 1961, pp. 225 ff.), congeniality between "ringer" and "bundler" was found to vary with status congruence: Congeniality was high when the ringer was equal to, or higher than, the bundler in both general work responsibilities and social position. Exline and Ziller (1959) set up congruent and noncongruent decision-making groups in terms of status derived from each member's ability on a task and the worth of each member's vote. Congruent groups were rated by observers as significantly more congenial. An investigation by Raven and French (1958) is also relevant. They predicted that the more a supervisor was perceived as having a legitimate right to her position, the higher would be the ratings she received from group members on a liking scale. The results, although not statistically reliable, were in the predicted direction. Judging status along a standard of legitimacy may be considered of the same nature as comparing status on one dimension with that on another.

Behavior or Personality Characteristics

Here we will consider investigations which have related the attractiveness of an individual to his specific characteristics, regardless of those of the individual attracted to him or the nature of their relationship.

One class of such studies is represented by those in which persons differentiated on the basis of personality test scores are compared with respect to their likability or the extent to which they are chosen by others as friends (or on some related criterion). Masling et al. (1955), in two separate investigations with army men, tested the hypothesis that "equalitarians" will receive more sociometric nominations than authoritarians because they exhibit greater warmth and less hostility. This assumption was based on previous research by Sanford and by Eager and Smith (cited by Masling et al., 1955) in which equalitarians were found to be perceived as warmer than authoritarians, and camp counselors judged as hostile scored higher on a measure of authoritarianism than those not so judged. Masling and his co-workers found that a small but significant positive relationship between equalitarianism and sociometric choice did exist where army status was held constant. Chapman and Campbell (1957) have also reported a small, insignificant correlation in the same direction: highly authoritarian individuals (as measured by the California F Scale) tended not to be chosen as desirable teammates. In a different kind of study, Kates (1959) had subjects rate two stimulus persons described in case studies. The objectively more authoritarian person was perceived by the subjects as having greater power, exhibiting more leadership and being more personally attractive. He was not, however, judged as being more authoritarian. These findings, then, do not contradict those cited above since the difference in authoritarianism between the two stimulus persons was not recognized by the subjects.

A number of investigators have pursued the question of whether individuals with relatively positive self-concepts are more attractive or better liked than persons with less positive self-concepts. McIntyre (1952) found no significant difference, in responses to a self-attitudes questionnaire devised by Phillips, between the most highly and the least accepted men within a college sample. A similar study by Fey (1955), using his own self-attitudes scale, also obtained no significant relationship between self-acceptance and acceptance by others, but subjects with high self-acceptance and low acceptance of others tended to be rejected by others. Positive findings with respect to self-acceptance and acceptance by others have been reported by Miyamoto and Dornbusch (1956) as well as by Reese (1961). In the latter study, of elementary school children, evidence was obtained for a curvilinear relationship, that is, high acceptance, measured by both rating and ranking, was found to be accompanied by moderate self-concept scores as measured by a scale devised by Lipsett.

Still other characteristics have been found to distinguish attractive, likeable individuals. Kelley (1950) had students in an economics class rate a substitute teacher, about whom they had previously received information, on seven personality-behavior scales, following a discussion led by the substitute. Those students to whom the substitute had been previously described as "warm" consistently rated him more favorably than those to whom he had been described as "cold." Using multiple criteria to determine acceptance and rejection among dormitory members at a girls' college, Lemann and Solomon (1952) found that highly accepted girls were rated by their peers as "generous, enthusiastic, and affectionate, while girls with low status were generally rated more stingy, apathetic, and cold." R. L. French (1951) obtained data among naval recruits which indicated that men with sick bay and disciplinary offense records tended to be less acceptable to their peers than were other men; and Tagiuri (1952), in comparing 15 maladjusted prep school boys (i.e., boys who had been seen by the school psychiatrist) with 15 well-adjusted boys, found that the latter received more positive sociometric choices from their peers than did the former. A similar phenomenon in a quite different environment has been reported by Speroff and Kerr (1952): Among a group of workers in a steel mill, interpersonal desirability was negatively correlated with number of personal accidents.

With the exception of Kelley's (1950)

study, the other investigations mentioned in the preceding paragraph leave unanswered the question of which variable, positive behavior or relative attractiveness, is antecedent and which is consequent. This is true, as well, of a study by Van Zelst (1951) in which construction workers who were rated high in desirability by others were found to differ significantly from those rated low, in being generally better adjusted to their jobs and more favorable in their attitudes toward the company. Causality is more apparent in findings reported by Kipnis (1961): Male freshmen who had ascribed negative traits to their best friends (more than to themselves) broke off these friendships, within a six-week period, more frequently than did those who had ascribed more positive traits to their friends than to themselves.

Pepitone and Sherberg (1957) found that subjects who read scripts of an incident involving a punishing person rated him significantly higher on a liking scale, that is, disliked him less, when he was perceived as having "good intentions" than when his intentions were believed to be bad. In another study with Wallace (cited by Pepitone, 1958), subjects who failed test items under arbitrary conditions imposed by the experimenter disliked him more than did subjects who failed justifiably. In both of these studies, an individual's attractiveness to others is seen to be influenced by his particular behavior in a given situation, regardless of whether this behavior is characteristic of his personality.

That both more enduring traits and situation-specific behavior determine attractiveness is suggested by the findings of Jennings' (1950) investigation of interpersonal relations among four hundred girls in a training school for delinquents. Girls who were overchosen by their peers on the criterion of "live and work with" are described by Jennings as sensitive "to the elements of the total group situation to a very much greater extent than the average member"; they contributed constructively to the development and positive experience of others, displayed high esprit de corps and tended to internalize private worries. Compared with the other girls, the highly chosen were less quarrelsome, less aggressive, more cooperative, and more initiative-taking. Similar findings have been reported by other investigators. For example, with respect to the variable of sensitivity, Gage and Suci (1951) found that high-school teachers who were more accurate in predicting their students' opinions on school related matters were also those who elicited positive affect from their students. Popularity has been related to accuracy in the perception of others' interpersonal preferences within infantry rifle squads (Greer, Galanter, & Nordlie, 1954), and to accuracy in the perception of others' popularity and group dimensions within classroom groups (Bugental & Lehner, 1958). Other characteristics of liked persons, similar to those reported by Jennings, are also supported by subsequent investigations. Haythorn (1953) found, among naval reserve officer trainees, that those men who were most often chosen by their peers on a number of sociometric criteria were those who most facilitated the effective functioning of their groups, who were rated cooperative, efficient, and insightful, and who scored high on personality measures of maturity, adaptability, and acceptance of others. Positive relationships have also been reported between likability and number of interactions initiated in discussion (Norfleet, 1948), judged value of contributions to the group (Theodorson, 1957), and frequency of participation (Shelley, 1960).

In the Theodorson study, the cited relationship was limited to certain groups, those high on an index of cohesiveness. Other limiting conditions are suggested by an investigation by Bales (cited by Riecken & Homans, 1954). Individuals in problem-solving groups rank ordered each other on four criteria at the end of each of several meetings. Scores on "contributed best ideas to solve problem" and "did most to guide discussion" correlated highly with frequency of both initiation and receipt of interaction, but "like" scores correlated highly with the other scores only at the first meeting and not at subsequent ones. A man who had ranked high at the end of the first meeting on both the "best ideas" and the "like" scales tended to keep the latter status and lose the former by the end of the second meeting. Thus, in certain kinds of groups, specification of which is still unclear, "task master" and "good fellow" roles (Riecken & Homans, 1954) may be antithetical. In another investigation, Bales (cited by Homans, 1961) found that group members tended to dislike those leaders who talked more than they were talked to while leaders "whose ratio of receipts to initiations was high were apt to be chosen highly for both guidance and liking [Homans, p. 306]." Related to this general question, Jennings (1950, p. 272) noted that although many of the choices made on the "live or work with" criterion were the same as those made for leisure time companions, there were also differences; and the foundations for the latter choices seemed to be "based more exclusively in warmth and good feeling, as ends in themselves."

A large number of investigations concerned with individual personality and small-group performance (reported between 1900 and 1957) have been reviewed by R. D. Mann (1959). He concluded that the data on popularity support a positive relationship with "extroversion, intelligence, adjustment, and conservatism."

It is apparent that the attractiveness of an individual is enhanced by certain behavior which may be either temporary and situation bound or enduring and central to his personality. This behavior, as seen from the research summarized above, can be described in terms of warmth, equalitarianism, good adjustment, sensitivity, helpfulness, and so on. Such behavior might be further described, more generally, as that which should result in interaction that would have pleasant, as opposed to unpleasant, consequences.

Similarity

No other variable considered thus far has been the object of more investigation than this one. One of the major problems, however, is that the question of whether mutual attraction is a consequence or an antecedent of similarity is not always answered, although there have been a few controlled studies directed specifically at clarifying the direction of the relationship. A related problem in this area of research concerns the measurement of similarity. B's similarity to A is sometimes objectively determined; other times A simply perceives B as similar to himself (and perhaps erroneously). In the latter case, does A perceive B as similar to himself because he likes him, or is he attracted to him because he believes that B is similar? It is probable that both relationships hold, but here we shall be concerned with the latter. Data bearing on the former will be considered later when we examine findings relevant to the consequents of attraction between persons.

Background (Race, Ethnicity, Occupation, Age, etc.). In a laboratory group composed half of Jewish girls and half of Catholic girls, the girls in each category split their votes equally between Jews and Catholics before members were identified by religious affiliation. After identification, the Jewish girls continued to split their votes, but the Catholic girls overchose within their subgroup (Festinger, 1950b). In a subsequent situation where religious identification of the *voter* was not possible, both Jewish and Catholic girls voted in the direction of their respective subgroups. Choice on the basis of similar religion has also been reported by Goodnow and Tagiuri (1952) among boys attending a liberal preparatory school; and Zaleznik, Christensen, and Roethlisberger (cited by Homans, 1961, pp. 214 ff.) found that in a Boston industrial plant persons in the Irish majority were much more likely to be approved and accepted than those of a different ethnic background. Other investigators have reported similarity of occupation to be a basis for friendship choice among adult members of training groups (J. R. P. French, 1951), and similarity in the education and salary of fathers to be a basis for not desiring a change in roommate among freshman girls (Broxton, 1962).

In a field study by Seashore (1954), the cohesiveness of industrial work groups, measured by responses to questions about the group as a whole, was found to be positively related to within-group age variance, not age homogeneity, and not at all related to similarity in educational level of members. Cohesiveness did correlate positively and significantly, however, with the percentage of members in the group who had had over three years of service with the company. It is not surprising to find that the importance of similarity on particular attributes will vary in different group situations, or, in other words, that homogeneity with respect to a given characteristic will not be related to friendship choices in all groups. Thus, Lazarsfeld and Merton (1954) have reported that within each of two housing communities, Craftown in New Jersey and Hilltown in Pennsylvania,

> the degree of similarity . . . of close friends varied greatly for different attributes . . . from the almost complete limitation of intimate friendships among those of the same race and sex, to entirely negligible selectivity in terms of educational status [p. 22].

Further, religion was an important selective factor in Hilltown, but not in Craftown, where similarity in political values took precedence.

What may be operating in the tendency to make positive choices for individuals similar to ourselves in background characteristics is suggested by the results of two studies reported by Byrne and Wong (1962). In the first, Wong had subjects indicate their feelings about strangers on whom only minimal background data were provided. As expected, highly prejudiced white subjects responded more negatively to a Negro than to a white stranger, while subjects low in prejudice did not. The second study tested the hypothesis that highly prejudiced white subjects assume a greater degree of *attitude dissimilarity* between themselves and an unknown Negro than between themselves and an unknown white person whereas subjects low in prejudice do not. The hypothesis was supported. Of further interest is the analysis of actual attitude similarity; there were not significant differences between a sample of white and Negro subjects on 26 items, indicating that the dissimilarity highly prejudiced white subjects perceived between themselves and Negroes was a function of their prejudice. A question is raised by these findings which merits investigation: If we tend to believe that persons who are like ourselves in some salient background characteristic are also similar to us in attitudes, will it also be the case that when A perceives B to share a number of his attitudes that he will assume B to be similar to himself in background?

Attitudes. In a further study by Byrne and Wong (1962), white subjects were provided with both background information and the results of a 26-item attitude questionnaire (dealing with a range of issues from God to television) for either a white or Negro stranger and asked to indicate their feelings toward (i.e., liking of) the stranger by checking one of seven statements. Half of the subjects were led to believe that the stranger's attitudes were in complete agreement with their own in direction, while the other subjects were led to believe that the stranger's attitudes were opposite to theirs. Personal feelings toward the stranger were found to be significantly influenced only by attitude similarity and not by race of the stranger, or by the subject's degree of prejudice. These data and other findings reported by Rokeach (1960) indicate that belief congruence may, under certain conditions, be more important than race in determining interpersonal preference. In this connection, Triandis (1961) has pointed out that whereas belief similarity may be more important than ethnic similarity with respect to friendship choices, this is not the case with respect to relationships "involving relatively large social distance, such as acceptance of a person as a neighbor or as a student in one's university."

Similarity in interest was found to be more closely associated with clique membership than either grade or neighborhood in a study of adolescents in a California high school (Marks, 1959), although the interests which were relatively homogeneous within cliques differed with the sex of the group. From this study it is not possible to determine clearly whether common interests preceded clique formation or whether membership in the same group gave rise to shared interests. In some other investigations, however, there is less ambiguity regarding the independent and dependent variables. Broxton (1962), for example, found that freshman women who, at the end of five weeks, did not desire to change roommates were more similar to their roommates in certain attitudes, for example, toward drinking and smoking, than were women who did want to change.

A considerable amount of relevant data has been reported by Newcomb (1958, 1961) from a large-scale study of university men, initially strangers, who were housed together in a dormitory. Measures of attitudes toward 100 objects, as diverse as President Eisenhower, house rules, and classical music, were obtained prior to the boys' acquaintance with one another. For the first of two groups of seventeen students, "it turned out that clusters of most closely agreeing individuals, before acquaintance, were those most likely later to constitute cohesive cliques [Newcomb, 1958]." The preacquaintance indices of actual agreement did not predict to early sociometric choices, but to those made after the first two or three weeks, after, presumably, the men had had a chance to learn one another's orientations to various issues. The attitudes of the men remained, in general, relatively stable while their "high attraction preferences tended to change in favor of individuals with whom they were more closely in agreement [Newcomb, 1961, p. 254]." During the early weeks, judgments made by the men about each other's attitudes were related to their feelings of attraction, but their judgments tended to be not very accurate. "With the acquisition of new information about each other, estimates of others' orientations tended to become more accurate, and the preponderance of changes in attraction were thus influenced by considerations of reality [1961, p. 255]."

Byrne (1961b, 1961c) has supplied additional data relating attitude-scale agreement with liking and has also reported (1961c) that the effect attitude similarity has on attraction varies with strength of affiliation motivation. Subjects high in need for affiliation responded to a dissimilar stranger with dislike but subjects low in affiliation need responded with indifference. It can be expected that attitude similarity will not be given the same weight by all persons when they judge the attractiveness or likability of another individual, and further, that similarity on some attitudes will be more important for some people, that is, have more influence on such judgments, than similarity on others. Supporting data with respect to the latter expectation have been reported by Newcomb (1961), but both hypotheses are in need of further verification and refinement.

Some studies have shown a relationship between attraction and agreement on a variety of relatively specific issues, as distinct from general attitudes. Festinger, Schachter, and Back (1950) found that in a student housing project consisting of U-shaped courts there was relative homogeneity within courts with respect to attitudes toward a tenants' organization, and that those who deviated from the dominant attitude tended to be rejected on the sociometric question, "Who do you see most of, socially?" In another part of the housing area, however, consisting of rows of two-story buildings, the degree of homogeneity regarding attitudes toward the tenants' organization was even greater than within the courts, but deviates were not rejected. In a theoretical paper, Festinger (1954) reported data from a study by Festinger, Gerard, Hymovitch, Kelley, and Raven (1952) which had not been given in the original research paper. Groups of subjects had studied a labor dispute and then evaluated the union's behavior. After the experiment, the subjects had been asked how well they liked the other persons in their group. "In each of the eight different experimental conditions those who thought that the others held divergent opinions were less attracted to the group [Festinger, 1954]." Gross (1954) found that for some Air Force personnel interpersonal attraction was associated with common satisfaction with the Air Force, or agreement with respect to personal commitment to Air Force goals, while for other men attraction was associated with shared dissatisfaction with the air site and with their jobs. In Hilltown, one of the housing communities discussed by Lazarsfeld and Merton (1954), white residents tended to overselect as friends those with the same "racial attitudes." Newcomb (1953), similarly, has reported that in two replications of a study in a small college community, students at each extreme with respect to political attitudes chose, as friends, those like themselves. Altman and McGinnies (1960) have related similarity in ethnocentrism among group members to their attraction to one another. Six-man groups, of varied composition with respect to the California E-Scale scores of the members, viewed and discussed a film dealing with ethnic minorities and prejudice. Low-ethnocentric (E) subjects were found to be more apt to choose other low-E subjects on an attraction question while high-E subjects divided their choices about equally between highs and lows. High-E subjects, however, were not as accurate as were low-E subjects in identifying the attitudes of other group members.

One study (Gage & Exline, 1953) has produced negative findings: no association between leisure-time, sociometric ratings of group members and either their actual or judged similarity in response to a series of opinion items regarding group processes and the study of human relations. Perhaps opinions on these questions were not especially important to the group members and a different set of items might have yielded different results. Newcomb's data (1961) suggest that agreement on some issues is more related to attraction than agreement on others, and that these are the issues about which the individuals are more concerned.

Provocative data within a different kind of research setting have been reported by Gerard (1961). Subjects in a modified Asch situation ranked highest, on a liking scale, the subject who always agreed with them in line judgments, and they ranked lowest the subject who never agreed with them. In addition, a group member who switched from disagreement to agreement either very early or very late in the judgment series was ranked higher than one who switched somewhere in the middle. In another investigation of an Asch group-pressure situation, Malof and Lott (1962), though they obtained no data on liking, found that the support of a Negro confederate who switched to making correct judgments was accepted by highly prejudiced white subjects (exposed to erroneous judgments by a white majority) almost as readily as it was accepted by white subjects low in prejudice. Acceptance of support was defined by a change, on the part of the critical subject, from incorrect to correct judgments following such a change by the confederate.

Agreement between Individuals A and B, not on Issue X, but on their evaluation of another individual, C, has also been shown to relate positively to A's and B's evaluations of one another. Newcomb (1956) found that this relationship increases over time among a group of students living in the same dormitory, and Davol (1959) reported that among men in a Veterans' Administration domicile those who either liked or disliked another man in common tended to like each other.

If B's agreement with A on C enhances A's attraction to B, we can further suppose that B's agreement with A on A, that is, agreement with A's self-evaluation, should have the same result. Newcomb (1956) found this to be the case with respect to agreement on both positive and negative traits. Backman and Secord (1962), similarly, found that girls in a sorority house both like and interact most with those girls whom they believe see them as they see themselves (with regard to selected personality traits), and they like least and interact least with those girls whom they believe to have images of themselves at variance with their own. These data were unrelated to differences in the social desirability of the traits.

Deutsch and Solomon (1959) investigated the consequences of agreement on evaluation of task performance. Female telephone-company employees were led to believe that they had performed poorly or well on a task and then received either an unfavorable or a favorable note from one member of their group. Although there was a clear tendency for subjects to respond more positively to a favorable note writer under all conditions, when this was held constant it was found that a note writer was more favorably evaluated when her opinion of a subject's task performance was consistent with the subject's own than when the note writer's opinion was inconsistent with the subject's. A further study (Howard & Berkowitz, 1958), has shown that agreement among persons in their evaluation of an individual's performance also affects his attraction to them. Subjects whose task performance was judged by four observers tended to be less desirous of working with the one whose judgment deviated from that of the other three, regardless of whether the deviant evaluation was positive or negative.

A final group of studies relevant to the present discussion are those dealing with the relationship between acceptance of A by B and A's consequent attraction to B which we reviewed earlier. We assume that, in most cases, acceptance by another is congruent with own feelings toward ourselves, that is, that self-attitudes tend to be favorable.

Values. The Allport-Vernon Study of Values (or one of its revisions) has been employed in a number of investigations to compare friends with nonfriends or persons who like each other more with persons who like each other less on similarity of value profiles. Newcomb (1956), for example, has reported a significant relationship between actual similarity in Allport-Vernon scores and attraction and also between estimates of agreement (i.e., subjects' perceptions of how others would rank order the six values in the scale compared with their own rankings) and liking (Newcomb, 1958). This latter relationship was greater after fourteen weeks of interaction among men in a dormitory than it was after two weeks. A. J. Smith (1957), concerned with the causal direction of this relationship, gave to each of his subjects two partially completed Allport-Vernon booklets, one containing responses identical to those previously made by the subject and the other containing dissimilar responses. Subjects then rated each of the individuals, whose Allport-Vernon booklets they had presumably seen, on their desirability as social companions and as work partners. Acceptance on both measures was significantly greater for the hypothetical person with similar values. Reilly, Commins, and Stefic (1960) have reported a slight tendency in the direction of greater similarity in value scores between friends than between nonfriends. In still another investigation utilizing the Study of Values, Jones and Dougherty (1959) found that subjects with high political scores evaluated another politically oriented person favorably, especially when competitive interaction with him was anticipated. When no interaction was anticipated, subjects with a high political value score chose as "liking best" a political rather than an aesthetic person. When, however, scores on Christie's Mach IV Scale were used as a measure of political values, contrary findings were obtained. Individuals high on the Mach Scale, who can be described as manifesting a strong need to manipulate the social environment, tended to devalue others with a similar orientation. These findings anticipate others to be discussed below which indicate that either similarity or complementarity between persons may function to increase their attraction to one another, depending upon the nature of the characteristic (value or personality trait) and upon a number of other, as yet unspecified, conditions.

A nonsignificant correlation between value homogeneity and the ratio of in-group to out-group choices made by members of natural college groups was reported by Eisman (1959) and also by Ramuz-Nienhuis and van Bergen (1960) who replicated the Eisman study in Amsterdam. The later investigators noted, however, that an in-group to out-group choice ratio may "not reflect personal attraction very accurately," since feelings about other persons outside the group "will affect the choice ratio, while it has nothing to do with the feelings of personal attraction toward the group members."

Precker, in two studies (1952, 1953), found that students tended to choose peer group members for association after college, and also to choose as college advisors individuals whose rankings of 39 value categories were similar to their own, without knowledge of the rankings made by the others. The correlation between rankings was highest for mutual student choices. Data reported by Thompson and Nishimura (1952) are also relevant here if we can assume that a person's conception of the "ideal personality" is a reflection of his system of values. Each individual within a sample of eight pairs of best friends filled out a 100-item personality schedule under a number of different instructions, including rating the traits of ideal and friend. Significant positive correlations were obtained between the ideal personality ratings made by pairs of friends and also between the ratings of ideal and of friend by the same individuals. One further study seems pertinent to the present discussion. Thibaut and Kelley (1959) have suggested that it is costly (in terms of effort and expense) to maintain friendships over a distance and that such relationships would, therefore, "be expected to show relatively high value similarity." In support of this hypothesis, they cited data (p. 41) obtained by Williams and others to the effect that greater agreement was found between a sample of individuals and their friends residing outside their immediate community than between the same individuals and their friends living close by.

Compatible Personality Traits. What constitutes compatibility has not, thus far, been clearly defined. Within the empirical literature there are numerous contradictory findings, but these apparent contradictions may simply reflect the fact that compatibility requires similarity in certain characteristics and *complementarity* in others. In other words, compatibility seems not to be an exclusive function of one variable or the other, but a complex function of both.

Some studies have shown a simple positive relationship between attraction and similarity of objectively measured personality traits. Lindzey and Urdan (1954) reported that pairs of students who chose to room together were more alike on personality measures than were those who rejected one another. Similarly, Izard (1960a), using Edwards' Personal Preference Schedule, found pairs of friends to be significantly more alike in profile scores than randomly established pairs. Izard also obtained significant positive correlations between friends, but not between random pairs on the specific traits of exhibitionism, deference, and endurance. In another investigation Izard (1960b) demonstrated that the similarity in personality profiles between friends exists prior to their acquaintance.

In a different kind of study, Cohen (1956) paired subjects on the basis of similarity or dissimilarity in ego defenses, or preference for projection. Following discussion of defense-provoking stories, the similar pairs were more attracted to one another than the dissimilar pairs while the projection-preferring pairs were least attracted to one another. These data suggest that sheer similarity in certain personality characteristics may not be conducive to satisfying interpersonal relations, and that compatibility between the traits of one individual and those of the other must be more broadly defined. Winch (1955) and Winch, Ktsanes, and Ktsanes (1955), for example, who compared husband-wife pairs with randomly matched men and women on need patterns (as measured by the Thematic Apperception Test and interviews) found the existence of complementarity among the married pairs, especially with respect to the assertive-receptive dimension. In another research context, Gross (1956), studying informal groups within an Air Force population, reported that some groups were composed of men with dissimilar characteristics and that in these "symbiotic" groups each man seemed to have something to offer the other; other groups (consensual) were made up of men with similar characteristics who shared a goal or a set of values. Schutz (1958, 1961) has more explicitly defined and investigated these two kinds of compatible groups, which he called reciprocal when there was complementarity of three basic interpersonal needs, and interchange when there was similarity in interpersonal needs. He has shown that such groups tend to be productive and cohesive and that their members tend to like one another and to desire continued contact.

Negative evidence with respect to both similarity and complementarity has also been reported. Thus, Bowerman and Day (1956), using Edwards' Personal Preference Schedule, found no support for either homogamy or complementarity in the needs of 60 college couples, regularly dating or engaged; and Hoffman (1958) found no greater inter-member preferences within groups that were homogeneous with respect to profiles on the Guilford-Zimmerman Temperament Survey than within heterogeneous groups. A confounding variable in the latter study, however, was the variation among groups in successful problem solution.

It has also been shown that under certain circumstances a person may choose to work on a particular task with a dissimilar fellow. Lerner and Becker (1962) demonstrated that a similar person is preferred for interaction in a mutual-gain or noncompetitive situation, while a dissimilar person is preferred for competition. No data were gathered on liking, but the results seem to suggest that the similar individual is the more likable since we prefer to share rewards with him. A related investigation was conducted by Israel in 1956 (cited by Homans, 1961).

Each subject saw both his score, on a series of logic tests, and that of the other members of his group. Each subject believed his score to be average while those of his fellow members varied from high to low. Subjects were then asked to rank order the others in his group with respect to their continuance in the group for further solution of logical problems. Under a condition where reward would be based on the group's total score, choices tended to go to high scorers; under a condition where reward would be based on score homogeneity among group members, choices tended to go to other average scorers; and under a condition where reward would depend on individual performance, choices tended to be directed toward low scorers. Again, these results do not indicate liking, but suggest, simply, that work-partner preferences will be made in such a manner as to maximize personal gain, although it was also the case that, overall, high scorers were most preferred.

A number of studies have been primarily concerned with the consequences for interpersonal liking of similarity not in the actual personality traits of individuals, as measured by some relatively objective instrument, but in the traits which they judge each other to have. Most notably identified with such research on "assumed similarity" is Fiedler. In an early investigation (Fiedler, Warrington, & Blaisdell, 1952), each of a group of fraternity members was asked to describe, by Q-sorting 76 personality items, himself and his ideal self, and to predict how his best-liked and least-liked fellow members would describe themselves. The results indicated that best-liked persons were perceived as more similar to both self and ideal self than were least-liked persons. No significant difference in similarity was found, however, between actual self-descriptions made by subjects and best-liked choices, and self-descriptions made by subjects and least-liked choices. The results of other studies have also suggested that assumed similarity with a liked person may be greater than actual similarity. Preston, Peltz, Mudd, and Froscher (1952), for example, found that husbands and wives tend to rate each other on specific traits in the same way they rate themselves and that this tendency is greater for happy than for unhappy couples, but correlations between self-ratings were negligible. Similarly, Reilly et al. (1960), obtained no support for either similarity or complementarity in the actual needs of mutual friends, as measured by the Edwards test, but among 15 correlations of the actual scores of subjects on single personality needs and the scores they predicted for their friends, 10 were significant and positive.

Still other data can be cited which support a relationship between assumed similarity and attraction. Lundy, Katkovsky, Cromwell, and Shoemaker (1955) found that self-descriptions (on Minnesota Multiphasic Inventory items) were more similar to descriptions made of chosen persons than of rejected ones and, further, that the description of a best-liked person was significantly more similar to a subject's acceptable self (traits described for both self and ideal self) than to his unacceptable self, while the reverse was true of the description of a least-liked person. Finally, Kipnis (1961), from a study of changes in self-concepts, reported that a sample of male freshmen perceived smaller differences in personality between themselves and those whom they chose for a variety of friendly, interpersonal contacts than between themselves and those they liked least.

In some investigations, the personality traits of individual members have been determined by means of group judgments. Utilizing such a technique in three separate groups, Maisonneuve (1954) found that "people who are attributed similar profiles tend to mutually choose each other, or . . . people characterized in a similar way tend to associate." In a study by Zimmer (1956), a measure of the "behavior tendencies" of group members was obtained by pooling the rankings (on eight dimensions) of each other made by the members. A trend was found in the direction of greater disparity between the judged personalities of subjects and their choice of discordant (annoying) group members than between the judged personalities of subjects and those group members they considered harmonious. Still another variation in procedure and measurement can be found in an investigation by Stotland et al.

(1960). They asked each subject, following interaction with others on a task, to predict how the *majority* of the members in his group would respond to a 30-item personality questionnaire, and also to respond to it himself. When only subjects of no prior acquaintance were considered, those who indicated a high desire to get to know their fellow group members differed from those with low desire in perceiving the majority of the group as significantly more like themselves.

The findings reviewed above, relevant to the relationship between assumed or judged personality of others and attraction, do not, for the most part, provide a satisfactory answer to the question of causal direction, that is, whether A's perception of similarity between himself and B disposes him to find B attractive, or whether A tends to see in a likable B personality characteristics which are congruent with his own. We have assumed, from the context of the investigations reviewed in this section, that they bear on the former relationship; later on we will consider others which seem more clearly relevant to the latter.

When we put all the evidence together, there can be little doubt that individuals tend to prefer friendly associations with others who are compatible to themselves in interests, values or personality. The major problem in this area is specification of what constitutes compatibility in situations of varying characteristics.

Success–Reward

Shared Success or Failure. Typically, data concerned with this variable have been obtained from experiments in which the success or failure of small groups, working on a single task or series of problems, has been manipulated. Utilizing such a design, Shelley (1954) found that members of successful groups had a more favorable attitude toward their group than did members of groups experiencing failure. Steiner and Dodge (1956) in effect manipulated success and failure by interfering, in one condition, with messages sent by group members to one another via lights and buttons. In this condition ("perceptual inaccuracy"), incorrect messages were received, while in another condition there was no interference and hence "accuracy." When no specific rules had been provided by the experimenter for members to follow in performing their task (design reproduction on a checkerboard), the former condition was significantly associated with less task efficiency, with more criticism of the group, and with greater rejection of own group members for a new task. Working with groups of second- and fourth-grade children, Heber and Heber (1957) gave some groups a high score on an arithmetic test, some a low score, and some no score and then measured changes in the mean ratings group members gave one another on a social distance scale. Under the low score condition, the ratings decreased; under the success or neutral condition, they increased, with the effect showing greatest permanence after success. Pepitone and Kleiner (1957) introduced the threat of losing a game tournament to high-status (winning) teams. Where this threat was negligible, intermember attraction increased significantly more than within high-threat teams. Between low-status (losing) teams told they would probably win and those told they would probably lose, however, there was no difference in subsequent in-group choices. It is unfortunate that there are no data reported on the level of interpersonal attraction within the winning and losing teams after they had experienced success or failure and prior to the introduction of other variables.

In a study of attraction and personality similarity in which group success was not manipulated, Hoffman (1958) found that in homogeneous groups successful problem solving by the group was positively correlated with intermember attraction. More recently, Zander and Havelin (1960) investigated interpersonal liking as a function of variation in the task competence of trios within larger nine-man groups. Subjects in the most competent trios made significantly more work-partner choices and fewer rejections within their own subgroup than did subjects in the poorest trios. Similar findings obtained on a scale of general attitude toward the subgroup. (The perception of similarity among subgroup members on characteristics other than task competence was also manipulated and found to have a greater effect on the peer choices and group evaluations of subjects in the poor than in the successful trios.)

Both Wilson and Miller (1961) and Myers (1962) studied the effects of success in competitive situations and utilized, as measures of liking, personality ratings made by each member of every other member. In the first-mentioned investigation, winning was associated with a larger positive shift in the ratings given to their own teammates than was losing; in the second study, high-success teams increased significantly in esteem for their teammates, while esteem for teammates diminished in low-success teams. Phillips and D'Amico (1956) had earlier reported that within groups in which members competed *against each other*, where rewards were evenly distributed, there was either no change or an increment in within-group choices on a sociometric question, in contrast to groups in which rewards were not evenly distributed, where within-group choices decreased.

Both Deutsch (1959) and Zander, Stotland, and Wolfe (1960) have reported data indicating a positive relationship between group success and the attractiveness of the group as a whole; in the latter investigation, however, responses to a scale measuring attraction to members of the group did not differentiate between subjects in failing and successful groups. There have been other reports of equivocal or negative findings. Thibaut (1950) found that groups of boys who tried to obtain better treatment from the experimenter and failed significantly increased in cohesiveness (as measured by proportion of own group choices), while groups of boys who were successful in improving their status did not. This difference in heightened attractiveness of own group members held, however, only between the central (more popular) members and not between the peripheral members of unsuccessful and successful groups. In a number of other studies the experience of common failure has been shown to influence the choice of one failing subject for another on a variety of criteria. Shaw and Gilchrist (1955), for example, found that after a sequence of task failures, subjects chose among themselves for friendly association more than they chose successful individuals. The failing subjects, however, had experienced interaction only with each other. Berkowitz, Levy, and Harvey (1957) found that subjects who were highly task motivated and received unfavorable evaluations of their task performance subsequently rated each other as high in attractiveness (on both work partner and social companion criteria) as did highly motivated subjects who received favorable evaluations. Burnstein and McRae (1962) studied groups of four white subjects and one Negro confederate in a situation where group members, separated by vertical partitions, communicated by means of written messages to solve logical problems cooperatively. Following evaluation of a group's performance, highly prejudiced white subjects rated the Negro higher on a liking scale when the evaluation was poor than when it was good. For white subjects low in prejudice, the performance evaluations made no difference in their rating of one another on the liking scale under the different conditions.

Is the experience of common failure by members of a group interpretable as shared threat or stress emanating from a source external to the group? It would be helpful, relative to this question, to know whether subjects in unsuccessful groups view their common failure as arbitrarily imposed by the experimenter or as due to their own lack of ability. In the former case, but not in the latter, heightened interpersonal attraction would be expected. Pepitone and Wallace (cited by Pepitone, 1958) have shown that when subjects consider their failure arbitrary or unjustifiable they dislike the investigator more than otherwise, but no data were reported on attraction to other group members. We may make use of the Pepitone and Wallace findings, however, in inferring that arbitrariness of evaluation probably did not operate as a variable in the Burnstein and McRae study since there were no reliable differences between the favorably and poorly evaluated subjects in their ratings of the tests' "fairness" or in how much they "liked" the experimenter. It is possible that the Burnstein and McRae findings are specific to evaluations which prejudiced white persons make of Negroes, that is, Negroes are more attractive to such white persons when they are poorly evaluated than when they are favorably evaluated. Other data obtained in the same investigation suggest that this may be the case. Favorably evaluated subjects rated the communication skills of their fellow members significantly higher than did unfavorably evaluated subjects, *but* the Negro confederate was considered less skillful compared with the average group member by members of successful than by members of unsuccessful groups. Findings from another study (Katz & Cohen, 1962) also suggest that white persons may be less attracted to a skillful Negro regardless of their own effectiveness. Teams composed of one Negro and one white subject solved problems cooperatively. In one condition ("assertion training") the experimenter varied the information available to each partner so that only one could confidently propose the correct solution, but each partner was favored with the better information on half the problems. In the other condition ("no training") neither partner received sufficient information to permit him to propose the correct solution confidently. Following this experience the Negro subjects in both conditions expressed the same degree of preference for their partners (for a future experiment), but the white subjects who received "assertion training" preferred their Negro partners *less* than did those in the "no training" condition.

Equivocal findings with respect to the general proposition that interpersonal attraction will be greater in successful than in unsuccessful groups have been reported by Kleiner (1960). In group situations where one member was clearly responsible for improving group performance (or, more accurately, decreasing the "likelihood of failure") there was no significant increase in subjects' ratings of each other, but only in their ratings of the one member responsible for the improvement. On the other hand, in the condition where there was only a small reduction in likelihood of failure, group members tended to depreciate each other, apparently "holding each other responsible for the relatively small improvement."

It seems clear that members of successful groups tend to like one another more than do members of unsuccessful ones. It is also true, however, that under certain conditions intermember liking follows shared failure, especially, we suggest, where the failure is perceived as arbitrarily imposed by an external source.

Attraction to Source of Reward. As in the case of shared failure, shared success, too, may be seen as emanating from an external source in a more or less arbitrary fashion, or it may be perceived as due to the skill or other characteristics of one's fellow group members. Such a distinction could not be made unambiguously in the case of the investigations discussed in the preceding section. Here we will consider research which has measured the attitudes of reward recipients toward those who were more or less directly responsible for the satisfying state of affairs.

The study cited above by Kleiner (1960) bears directly on the proposition here considered. Groups of two subjects and one "stooge" worked on puzzles, and their ability to solve problems was compared with that of business executives. In each group the stooge was clearly instrumental in improving group performance. When the improvement produced a large reduction in threat of failure, consequent liking and desire to know the confederate were significantly greater than when the threat reduction was small. Attraction to the helpful stooge also tended to increase more in groups under initially high threat of failure than in groups under low threat. A somewhat similar design was used to test a related hypothesis by Ziller and Behringer (1960). Groups solved problems and were informed of their success or failure. Into each group, then, a new member was introduced and a more difficult problem (which he could solve) was presented. The newcomer was rated higher by subjects in previously failing groups than in previously successful ones; it was in the former groups, of course, that his presence was associated with the greater improvement in group performance.

Earlier, Jennings (1950) reported that girls highly chosen by others to live or work with were sources of need gratification, and she suggested that choice on such criteria "may, in a sense, be considered 'earned' by the person chosen [p. 165]." Horwitz (1958) found that in an experimental simulation of a classroom situation the teacher was significantly more liked and was evaluated more positively on personal characteristics when he behaved in accord with the students' desires to repeat parts of a lesson than when he behaved in accord with his own desire, contrary to that of the students. In a different kind of situation (Solomon, 1960), subjects who communicated with unseen partners via electric switches developed greater liking for those partners whose game strategy provided them with maximum gain. Also reported have been data supporting a relationship between the feeling that one has influenced others (which we can assume to be gratifying or rewarding) and the liking of these others (Zander & Cohen, 1955; Tagiuri & Kogan, 1960).

Two studies have indicated that behavior by one person which is generally considered rewarding to another will not be so under all conditions. Howard and Berkowitz (1958) have reported that when judges disagree in their evaluations of a person's performance, the deviant judge tends to be seen as probably in error and tends to be rejected on a sociometric question regardless of whether his evaluation was positive or negative. Related findings were reported by Deutsch and Solomon (1959). Although subjects in their experiment tended, generally, to respond more favorably to persons who evaluated their task performance positively, they were less favorable if they had been led to believe by another source that their performance was poor, and under this condition they responded favorably to a negative evaluator. It appears as if a "correct" judgment may constitute desired behavior in situations where individuals received feedback about their actions from more than one source and adopt either one or the majority evaluation as their standard.

Individual Reward in the Presence of Others. A different interpretation of the role of success in influencing liking between individuals is that A's attraction to B does not depend upon B's having also shared success within the same group nor upon B's instrumentality in obtaining reward for A; instead, B's presence when A experiences a satisfying state of affairs is hypothesized as being a sufficient condition for enhancing A's attraction to B. The first two conditions have not, unfortunately, always been clearly absent in the investigations we will consider below.

Bass (1955) had subjects in a group rank words according to high-school boys' familiarity with them, first privately, then after group discussion, and then again privately, followed by the correct ranking given by the experimenter. Each person's own success or failure was found to be a significant determiner of his attraction to the entire group. Group goal attainment in this study was unrelated to individual attainment which was primarily a function of individual competence. Similar data have been reported by Spector (1956). Among men operating in the laboratory as military intelligence decoding teams, those "promoted" to sergeant judged their team as being significantly more attractive, and desired to remain within it, more often than those not promoted. Further, among the promoted subjects those who believed the probability of promotion to be low were even more favorable in judgments of their group than were those who believed the probability of promotion to be high. Although it is not clear what the achievement of higher status would depend upon, it is doubtful that the subjects believed their teammates to have been instrumental in their promotion or nonpromotion. An experiment by Lott and Lott (1960) was specifically designed to test the hypothesis under present discussion. Three-member groups of children played a game in which some members were successful and others were not. Later, on a sociometric test given outside the game situation, the successful children chose a reliably greater proportion of their fellow group members than did the unsuccessful children. Whether the former believed that their fellow members had contributed to their success is a question on which, unfortunately, no data were obtained. James and Lott (1964) extended the above experiment by varying the frequency of reward obtained in the presence of others and by using an additional test of attraction to group members. On each of the tests significantly more of the subjects who received six rewards chose fellow members than did those who received three or no rewards; no reliable differences were found between the latter two conditions. In both this and the previous study the number of rewarded subjects within each group varied, and later examination of the data has revealed some tendency for rewarded subjects to prefer other rewarded subjects.

Indications that certain personality factors may affect the influence which one's own success or failure has on our affective reaction to others may be found in data reported by Stotland (1959). Members of small groups first worked together, then separately. In the latter condition some failed and others succeeded on either a group relevant or irrelevant task. Attraction to the group was then measured by two questions: Would you like to meet with the group again? How much do you like the other members? Success on the individual tasks was associated with reliably more attraction to the group than was failure, but among those who failed, subjects with high self-esteem found their group more attractive than subjects with low self-esteem. There was no reliable difference between high- and low-esteem subjects who succeeded.

Negative findings have been reported by Rosenbaum (1959) who found no reliable difference in responses to a social distance scale (regarding their partners) between subjects who failed and those who succeeded on a discussion task. Instead subjects who were given no feedback about their performance liked their partners significantly more than did those given either favorable or unfavorable evaluations. It may be important that half the subjects were told that both partners would fail or succeed together, while the other half were told that if one partner succeeded the other would fail. No measure of the interaction between this cooperation-competition variable and success-failure is given.

Attraction to Successful Persons. Gilchrist (1952) found that successful individuals tend to be chosen for further activity by both other successful and unsuccessful persons. While this kind of choice does not necessarily indicate "liking," Gilchrist also reported a generally greater choice of successful subjects for social interaction (to speak to at the next meeting, have coffee with, double date with). Berkowitz (1957) reported, as an incidental finding in a study designed to test another hypothesis, that subjects liked their partners (simulated) better when the latter were supposedly proficient on a task than when they were supposedly poor. In another investigation (Berkowitz et al., 1957) where the performance of group members was evaluated by "experts," highly task motivated subjects generally indicated greater attraction to favorably evaluated than to unfavorably evaluated group members on both work partner and social companion questions. In still another relevant study by Stotland and Hillmer (1962), similar data were obtained: Subjects tended to show a greater desire to converse with a hypothetical person they were told was good at a task than with one they were told was poor.

What the effect one's own success or failure may have on the tendency to find a successful other attractive is an important question on which empirical attention needs to be directly focused. Existing data have not clarified the nature of the interaction between the variables of own goal attainment and others' in affecting interpersonal attraction. It is probable also that the interaction will vary from one set of circumstances to another. We have already seen that in certain cases unsuccessful persons show a preference for others who are "in the same boat," but the general nature of conditions under which this occurs is far from clear. We have also seen that when a white person and a Negro are equally competent on a task, the former tends to find the latter less attractive (because it violates the white person's expectations or reduces the value of his own attainment?). It is possible that under still other conditions the variables of own and other's success may operate independently so that an individual will be more positively disposed toward the members of his group or toward persons in his immediate environment following his receipt of reward, but his attraction toward other rewarded persons will be even greater than that toward nonrewarded ones. Future research might profitably be directed toward clarifying such questions and focusing on the interaction of variables.

Reduction of Cognitive Dissonance

It has been suggested that one way of reducing dissonance relative to cognition of another person is by increasing one's attraction toward him. In such a case it must be assumed that alternative methods of reducing dissonance are either unavailable or less likely to occur for one reason or another.

Supporting data have come from an investigation by Aronson and Mills (1959). They predicted that persons who undergo an unpleasant initiation in order to gain admission to a group and then find that the group has not been worth such pains (a dissonant cognition) will consequently exaggerate the group's positive characteristics. An alternative way to reduce the dissonance by convincing oneself that the initiation was not very unpleasant was made unlikely by the use of a severe initiation condition involving the reading aloud of obscene words. University girls undergoing this procedure were expected to find the group into which they were accepted more attractive than girls who were accepted without undergoing an initiation or whose initiation was "mild." After the subjects heard a "very dull" taped discussion, presumably being carried on by the group, they rated both the discussion and the participants. Responses to both measures supported the hypotheses.

It may be important that in this experiment the subjects were *accepted* into the group. It is possible that acceptance after a severe initiation is more valuable, more rewarding, than acceptance after a mild or no initiation. A condition of no-acceptance following the initiation procedures, with the same dull discussion and its participants subsequently evaluated, should provide worthwhile data.

Schopler and Bateson (1962), to test another alternative explanation of the above findings, replicated the Aronson and Mills study under somewhat altered conditions of subject recruitment and the addition of two new experimental groups. They found that subjects in the severe initiation condition rated the content of the dull group discussion significantly higher than did subjects in a mild initiation condition (the control group), thus confirming Aronson and Mills. Subjects who received a severe initiation, however, and also heard the experimenter *disparage* the discussion, did not differ in their ratings of the discussion from the control subjects and made reliably lower ratings than did subjects in the severe condition. Working within the Tribaut and Kelley (1959) theoretical system, Schopler and Bateson explain their results by arguing that in the severe condition, high dependence upon the experimenter is produced and, thus, a greater necessity for subjects to increase their power:

> One of the likeliest ways of increasing power which was available . . . was to conform to the opinion of the *E* [experimenter], who strongly implied that the *S*s [subjects] would like the discussion [Schopler & Bateson, 1962, p. 637].

In the disparage condition, introduced by Schopler and Bateson, where the experimenter predicted to the subjects that the discussion would probably not be good or exciting, the subjects rated the discussion low. Another condition introduced by Schopler and Bateson yielded ambiguous results. Of more concern to us are the findings with respect to ratings of the discussion *participants*. Although severe condition subjects rated the participants higher than did either the control or disparage condition subjects, none of the differences were reliable.

ANTECEDENTS OF LIKING: THEORETICAL EXPECTATIONS

The foregoing review of research, organized in terms of empirically tested relationships between a variety of conditions and consequent interpersonal attraction, should, we believe, have demonstrated at least one point. If we define the cohesiveness of small groups in terms of the positive judgments which members make of one another, then a great deal is known about the conditions under which cohesiveness is likely to develop. We suggest that the conclusion of Cartwright and Zander (1953, p. 80; 1960, p. 79) that there is "very little systematic knowledge about the conditions which heighten cohesiveness" is not an accurate reflection of the present state of affairs, although perhaps we do best to remove the qualifying adjective "systematic." We must, similarly, question Shaw's (1961) statement that little sound research has been carried out in an attempt to investigate group development. We do share, with Shaw, the conclusion, that there is "a tendency to ignore much of the literature relevant to the problem under consideration . . . in both theoretical works and research reports." This state of affairs is no doubt related to the vastness of the existing literature as well as to its generally nontheoretical nature.

We turn now to a consideration of theory, to attempts which have been made, in other words, to predict specific relationships on the basis of a set of more general principles.

Balance or Congruity

Heider. Most clearly identified with the above position as explicitly related to the question of attraction between persons is Heider (1946, 1958). His most basic concept, that of the "balanced state," refers to a situation in which the perceived units (i.e., persons in a subjective relationship, or person and some object or idea) and the experienced sentiments (i.e., like or not like) "co-exist without stress [1958, p. 176]." He maintains that under most conditions there is a general tendency towards a balanced state, that this is the preferred state of affairs. From this general principle follow a number of specific predictions.

We will consider here only those predictions in which sentiment formation is the dependent variable: A person tends to like a similar individual (but not when the similarity carries with it disagreeable implications, as, for example, when the other individual has the same disability as the person); A person tends to like another with whom he has contact (but not if continued interaction yields the perception of dissimilarity with the other); A person tends to like a familiar other (but not if the other is dissimilar). In each of the above hypotheses it is B's similarity to A which underlies A's attraction to him. Further, if another individual's behavior is perceived as congruent with a person's values, the person will tend to like the other. Whether similarity is implied by "balance" or whether it has some *other* value is not clear. We may note Heider's statement that ". . . it is satisfying to find support for one's own view [1958, p. 196]," but we are not sure whether support for one's own view is seen as "satisfying" a person's need for balance; that is, whether because A relates to B they should be similar to one another to achieve a balanced state or whether support for one's view "satisfies" a different motive. Since balance is said to be achieved "if the relations between . . . entities are all positive [1958, p. 202]," it may be that Heider considers similarity to be intrinsically *positive*.

There are two other relevant hypotheses which are, at least superficially, unrelated to similarity: A person tends to like another whom he has benefited and also one who has benefited him. To be benefited by a person one dislikes or to benefit someone one dislikes would obviously be disharmonious or incongruous.

All of the above predictions have been supported to varying degrees by empirical data, although surprisingly few investigations have been explicitly designed to test Heider's system. More often, the hypotheses tested in relevant research have been unrelated to any systematic point of view or have stemmed from other positions. Heider's influence is apparent, however, in other theoretical formulations. Newcomb (1953), for example, has similarly built a system around the basic principle of balance, postulated a "strain toward symmetry," and stressed the role of similarity between persons as the major determinant of attraction. We interpret Newcomb's more recent position, however, as more relevant to a reward framework and will, therefore, return to it below.

Festinger. The increasingly influential theory of cognitive dissonance has been applied to the prediction of behavior in social contexts in a paper by Festinger and Aronson (1960). This theory, too, focuses on balance and rests upon the basic assumption that individuals who experience simultaneous cognitions which are contradictory or inconsistent with one another will tend to reduce this incongruity (dissonance) by changing one of the cognitions in the direction of consistency with the other.

Regarding the problem with which we are here concerned, Festinger and Aronson have suggested that liking for an individual or group of persons may be a dissonance reducing mechanism. When, for example, a person has invested time or effort in order to interact with other persons and then finds that they are less pleasant than he had anticipated, he is faced with dissonant cognitions.

> There are at least two ways that a person could reduce dissonance in such a situation: (*a*) he could undervalue the amount of his investment . . . or (*b*) he could overvalue the group by emphasizing its positive aspects . . . [p. 220]

We have previously considered evidence relevant to this hypothesis.

Reward

It is possible to find reinforcement propositions within the more recent theoretical statements of a great many social psychologists (e.g., Allport, 1962; Bass, 1961; Miller, 1963). It is, for example, taken for granted that individuals are attracted to groups as a direct function of the satisfaction they are able to derive within them. Our particular interest here is in more explicit propositions regarding attraction to individuals.

Lazarsfeld and Merton. Although these writers have stressed common values as the dominant factor in friendship formation, the variable of reward seems, nevertheless, to be more primary. "For those with similar values," Lazarsfeld and Merton (1954) suggest, "social contact, because it is rewarding, will motivate them to seek further contact." In some cases friendships will form before the individuals become aware that they differ on particular issues. If there have been sufficient prior rewards, that is, sufficient areas of similarity, then some disagreement can be tolerated, but the tendency will be to modify opinions and come closer together if the friendship is to endure, since "common values make social interaction a rewarding experience." They suggest that the most critical evidence with respect to this question might well be obtained from data on broken friendships.

Newcomb. Resembling the above in general outline, Newcomb's position is more detailed and refined. He has chosen to distinguish between attitudes toward objects and attitudes toward persons, using the term "attraction" to refer to the latter (Newcomb, 1956), and postulated that the reinforcement which one person experiences as he interacts with another is the major independent variable determining attraction to the other. Specifically, positive attraction is conceptualized in terms of reward-associated attitudes, and negative attraction in terms of punishment-associated attitudes (Newcomb, 1960). Thus, behind the relationship between propinquity and attraction, according to Newcomb (1956), is the fact that "when persons interact the reward-punishment ratio is more often such as to be reinforcing than extinguishing." In the same vein preference for persons resembling others whom we like is accounted for by the "principle of generalization," that is to say, thresholds for interaction with such persons are lower and "the likelihood of the rewards of interaction with such persons is greater [1956]" than for others.

Perhaps Newcomb's most unique contribution to systemization in this area has been his emphasis on the role played by communication in the maintenance of friendship. Communication is recognized as the major interaction process through which individuals can reward one another. In addition, accurate communication is, in itself, rewarding, and consequently "reward value will attach to the co-communicator—which is to say that positive attraction toward him will increase (other things equal) with frequency of accurate communication with him [1956]."

Similarity between persons, again because it is rewarding, is said to account

> for more of the variance in interpersonal attraction than does any other single variable [1956, p. 579].

Newcomb has refined this relationship by noting that similarity, to be an effective determiner of interpersonal attraction, must refer to attitudes toward valued objects, that is, those which are "important and relevant [1958]."

> An individual tends to be attracted toward others who are seen as viewing as important the same things that he himself regards as important, and as taking attitudes toward them that are similar to his own. Upon first acquaintance, persons are "scanned" for clues to their attitudes, but initial judgments are not likely to be very accurate about a very wide range of objects. At first, therefore, attraction should be more closely related to perceived than to actual agreement. Given ample opportunity for association and communication, judgments tend to become more accurate [1958, p. 188].

These predictions have been verified in Newcomb's (1961) later research.

From the above it is clear how Newcomb relates proximity to similarity and then to attraction. It is only to the degree that proximity "facilitates the development of perceived similarity of attitude [1956]" that it contributes to attraction. Communication is also related to similarity since the latter tends to result from the former, at least with regard to the object of communication, and both accurate communication and similarity are independent sources of reward (1958). In like manner, the fact that individuals are attracted to persons who reciprocate their feeling is seen as "a special case of perceived similarity," assuming that self-attitudes tend to be favorable.

Newcomb (1960) is also concerned with the categories of reward which are significant to interpersonal relations and has distinguished among *admiration* (A attributes likable characteristics to B), *reciprocation* (A perceives B as liking him), and perceived *support* (A attributes to B attitudes similar to his own). A's attraction to group members would then depend upon two major parameters: his "hierarchy of 'appetites' for reward" (his motives), and "the reward possibilities that other members have to offer."

A considerable body of data already provide support for the major predictions within Newcomb's theory. Most valuable for future research would, we believe, be exploration of the manner in which the parameters mentioned in the preceding paragraph function in determining attraction.

Homans. In his recent book, Homans (1961) has strongly embraced modern learning theory as a source of principles regarding social behavior. As compared with Newcomb, therefore, Homans' theoretical propositions are more explicitly phrased in behavioral terms, but both systems depend upon the basic concepts of reinforcement and punishment.

Homans is concerned with both activities (overt behavior) and sentiments (overt behaviors which provide visible signs of internal attitudes and feelings), but since the former concept includes the latter, no separate sets of propositions are required. A primary assumption is that the activities and sentiments which one individual emits in response to another are reinforcing or punishing to some degree or "more or less valuable" to the other (p. 34). Special stress is placed on social approval (a sentiment) as a generalized reinforcer. The value of a reward to a person will depend upon his relatively constant past experience (his genetic and cultural character) and his present state of deprivation. "Costs" are subtracted from reward to measure the individual's "profit," or net reward.

In Homans' terms,

> cohesiveness refers to the *values* of the different kinds of rewards available to members of the group: the more valuable to a group's members are the activities (or sentiments) they receive from other members or from the environment, the more cohesive it is [p. 88].

With regard to interpersonal attraction, the following major hypothesis is derived from the general assumptions: The more valuable a person's activities are to others, the greater is the esteem in which he is held (p. 164). From this, in turn, follow certain specific expectations, for example: Persons will provide more value to one another if they share the same orientations, or have similar backgrounds (p. 214 f.); Because some individuals provide the most value to others, mutual choice will be greatest among high esteem, or overchosen, persons (p. 164 f.); Individuals tend to prefer high status individuals because such individuals have most likely acquired their status by rewarding others, and hence they are associated with the expectation of reward (p. 314).

Homans has suggested, further, that men tend to value "fair exchange" or "distributive justice." This "is realized when the profit . . . of each man is directly proportional to his investments: such things as age, sex, seniority, or acquired skill [p. 264]." From this follows the specific expectation that a condition of status congruence among persons will be conducive to mutual esteem or congeniality.

The relationship between interaction and favorable sentiments is also explained on the basis of the value derived from the interaction:

> the chances are that each one [of an interacting pair, for example] will find some of the other's activities valuable, if only because they may be obtained at less cost from him than from a third party at a greater distance: the distance of a source of reward adds something to the cost of getting it [p. 183 f.].

Homans has considered the contradictory findings which have been obtained with respect to the liking of leaders. He points out that a man "earns authority by acquiring esteem, and he acquires esteem by rewarding others [p. 314]," but since a man holding authority is likely also to incur costs on others by use of punishment or simply by depriving others of the chance to be the leader, the sentiments which develop toward him are likely to be ambivalent (p. 299 f.).

Homan and Newcomb offer similar predictions from similar frameworks, but Homans' specific accounting of "costs" has permitted a more detailed explanation of certain phenomena. Investigations designed to test the hypotheses relating both rewards and costs to esteem, or liking, are needed.

Thibaut and Kelley. An analysis of social behavior strongly resembling that of Homans' has been published by Thibaut and Kelley (1959). Although the Homans' (1961) book which we have cited above appeared subsequent to that of Thibaut and Kelly, we are considering the respective systems in the present order because they seem to fall best into these positions on a continuum of similarity.

Like Homans, Thibaut and Kelley analyzed social interaction in terms of rewards (satisfactions or reduced drives) and costs (physical or mental anxiety which serves to inhibit performance). In addition, they introduced the concept of *control*, which is described as mediated by one's ability to affect the outcomes of another's behavior. Thus,

> the problem that confronts a number of people if a continuing relationship among them is to be viable . . . is defined in terms of the possibilities the participants have for contributing to or detracting from each other's outcomes [p. 4].

A unique feature of the Thibaut and Kelley approach is its distinction between two evaluative standards which persons apply to the outcomes, that is, to the rewards and costs, of interaction. One standard is in terms of what the person feels he deserves, based on his past experiences and present motivation (the *comparison level*), while the other is in terms of available alternatives and is the lowest level of outcomes which the person will accept (the *comparison level for alternatives*). A person's dependence upon a group (or another individual) is said to be a function of the latter, but his attraction to the group or other individual depends upon the former. The two evaluative standards, however, are not independent. "Since outcomes from the best alternative relationship are included among those that determine the CL [comparison level], the CL and CL_{alt} [comparison level for alternatives] will tend to be positively correlated [p. 102]," and so, then, will be dependence and attraction.

It is with attraction that we are concerned, and Thibaut and Kelley (p. 24) proposed that a "person's attraction to a dyad depends upon his evaluation of his outcomes in relation to his CL . . ." Specific relationships between antecedent variables and interpersonal attraction are explained in terms of both *"ability to reward* each other and the *cost of providing this reward."* Value similarity is important to friendship because "in many value areas an individual is in need of support"; agreement is therefore rewarding. "In other words, provision of opinion support may be considered as having learned reinforcement value [p. 42]." Where complementarity contributes to the rewards obtained from interaction, then it will be positively related to attraction (p. 47). Similarly, we tend to prefer high status people because they "have greater extrinsic means . . . for rewarding others," or association with such persons may simply provide the reward of "prestige [p. 48]."

Other Reinforcement Formulations. The theories of Homans, of Thibaut and Kelley, and, to a somewhat lesser extent, of Newcomb are concerned with the entire sphere of social behavior. We have selected only what we believed to be most relevant to the present problem. Other formulations have been more narrowly concerned only with attraction among persons.

Pepitone (1958), for example, maintains that the simple common assumption that attraction to an individual "depends upon the amount of 'reinforcement' or 'need satisfaction' obtained from that person . . ." is far too general to provide an understanding of determinants. "What we want to know is what particular factors reinforce, reduce tension, etc. . . ." He has suggested that important dimensions are those of responsibility (for an act), intentionality (i.e., good, bad, or none), and justifiability (on some logical, social, or ethical standard). According to Pepitone, each of these has a predictable relationship with attraction: A's attractiveness to B will vary with the degree of B's responsibility for positive or negative acts, with B's intentions, and with the degree of justifiability of the positive or negative acts. These factors vary the value to one individual of the specific behavior of another. They *weight* the value (valence) of acts, and it is the value of his acts which determines the attractiveness of the actor. Status is related to attraction, for example, via the mediation of the above constructs, since high status persons will be seen as more responsible for an improved state of affairs, as having good intentions, and as being justified in their behavior.

A number of empirical studies, referred to earlier in this paper, designed by Pepitone and his co-workers to test his predictions, have yielded generally positive results.

Hypotheses regarding attraction among group members have been derived from the general propositions of Hullian behavior theory by Lott (1961). Attraction is rephrased in terms of positive attitudes and defined, following Doob (1947), as implicit, anticipatory goal responses. On the basis of a sequence of learning theory principles, it is expected that stimuli associated with reinforcement can eventually evoke the implicit and anticipatory component of the response to reinforcement, the goal response. Hence, the "primary condition for the development of mutual positive attitudes among group members" is said to be "the attainment of goals or the receipt of rewards in one another's presence." Secondary hypotheses regarding antecedents take into account frequency of reward and delay of reward. It is also predicted that the more closely or the more frequently a particular group member is related to the achievement of goals, the greater will be the strength of the positive attitudes developed toward him. This latter hypothesis may be interpreted as relevant to status and leadership. Studies designed to test this formulation have been cited earlier; other investigations are in progress.

Overview. There is clear agreement among many contemporary theorists that attraction will follow if one individual either directly provides another with reward or need satisfaction, is perceived as potentially able to do so, or is otherwise associated with such a state of affairs. Furthermore, the specific antecedent variables which empirical research has shown to be related to interpersonal attraction can, for the most part, be interpreted in support of this general proposition. The data on propinquity, group climate (including cooperation, democratic leadership, etc.), acceptance, status, personality characteristics, similarity (or complementarity), and success are all predictable from a general reinforcement position. It is more difficult to deal with the research findings on mutual threat or failure. If it can be shown clearly, however, as some investigators have suggested, that mutual threat increases interpersonal attraction only under conditions where the possibility exists of a cooperative solution to the common problem, then the latter condition could be identified as the source of reward.

CONSEQUENCES OF LIKING: EMPIRICAL RELATIONSHIPS

The investigations which are considered in this section may be roughly divided into the following types:

1. studies in which natural, existing groups are first measured for the degree of interpersonal attraction among their members, and scores on this measure are then related to some other variable;
2. studies in which groups assumed to differ in level of liking among members (as for example, groups of friends and groups of strangers) or groups known to differ on some variable which is assumed to reflect interpersonal attraction are compared on some other variable or dependent measure;
3. experiments in which interpersonal liking is explicitly manipulated through instructions or some other operation, and its consequences for some dependent variable are subsequently ascertained;
4. experiments in which interpersonal liking is indirectly varied through the direct manipulation of some related condition, and measures are then taken on a dependent variable.

Expression of Aggression

In an early study, French (1941) observed groups of strangers and groups of team or club members in a frustrating situation in which the group attempted to solve insoluble problems. A variety of measures indicated greater expression of interpersonal aggression within the "organized" than within the stranger groups, without major disruption of group functioning. Similarly, Wright (1943) found that pairs of children who were strong friends expressed more aggression toward the experimenter than did pairs of weak friends when they were subjected to frustration. Pepitone and Reichling (1955) used instructions to create high- and low-cohesive groups. The instructions informed members of the former that they would make "an exceptional team" since they were well matched, compatible, and would get along well; opposite expectations were presented to members of low-cohesive groups. Following provocation by a rude and insulting assistant, high-cohesive pairs were found to express greater hostility than low-cohesive pairs.

It has been suggested by Berkowitz (1958) that liking among group members has the effect of lowering inhibitions with regard to aggressive behavior. He has shown, however, in two investigations (Berkowitz & Green, 1962; Berkowitz & Holmes, 1960) that under conditions where aggression is displaced onto a person other than the instigator, generalization of aggression will vary positively with the degree of dislike of the person. In the first mentioned study, subjects who had been frustrated and insulted by the experimenter administered more shocks to disliked than to liked partners. In the second experiment, subjects subjected to similar treatment evaluated disliked partners on an adjective check list more unfavorably than they evaluated liked partners. Is it the case that a person will more freely direct hostility to the source of his frustration, that is, to an outside instigator, when he is in the presence of others whom he likes, but that he will tend not to displace aggression onto liked persons when disliked persons are as readily available?

Self-Evaluation

Manis (1955) has shown that, over time, there is a greater increase in agreement between the self-concepts of a sample of male freshmen and their friends' perceptions of them than between self-concepts and their nonfriends' perceptions of them. Two studies (Maehr, Mensing, & Nafzger, 1962; Videbeck, 1960) varied the evaluation made of subjects' performance by "experts." Assuming a tendency to be positively attracted to an expert, the results of these investigations may be considered relevant here: self-concepts were found to be positively influenced by both approval and disapproval. Thus, there appears to be some evidence that an individual will tend to change his self-evaluation in the direction of achieving congruence with the evaluation made of him by a liked other.

Negative evidence, however, has recently been reported by Pilisuk (1962). Subjects tended to retain favorable estimates of their own performance in the face of adverse criticism, whether the criticism came from a friend or from a stranger. When the criticism was alleged to have come from a friend, however, the subjects made use of a wide range of rationalizations to explain their friends' behavior. Harvey, Kelley, and Shapiro (1957) and Harvey (1962) exposed subjects to gradations of fictitiously negative information about themselves and found, similarly, that whether this information came from a friend or stranger made no significant difference in subsequent self-ratings. Subjects in the latter investigation, however, attempted to explain friends' derogatory evaluations differently from strangers' and tended to reevaluate friends more negatively than they reevaluated strangers who had rated them unfavorably. In the earlier study (Harvey et al., 1957), on the other hand, it was found that strangers who had made negative evaluations tended to be subsequently devalued whereas distortions were made of the negative ratings themselves when these presumably came from friends. It is clear, then, that criticism by a friend does not invariably lead to changes in self-evaluation. Whether negative evaluation leads to greater attempts at rationalization or devaluation of the source or the use of different distorting mechanisms when the source of the criticism is a friend in contrast to a nonfriend is not clear.

Self-evaluation was found to be influenced by persons' evaluations of their friends in a study by Kipnis (1961). She compared self-evaluations of male freshmen with their evaluations of others and found that the subjects tended to describe themselves, after a six week period, in the same way that they had previously described their best friends. Subjects who had ascribed relatively "good" traits to their best friends as compared with themselves changed their self-evaluations in a favorable direction, and those who had ascribed "poor" traits to their best friends changed their self-descriptions in a negative direction.

Another small group of investigations has been concerned with evaluation of own performance as a function of the performance or ability of liked others. Data reported by Rasmussen and Zander (1954) from a field study of secondary school teachers indicated that those who evaluated their own teaching performance in a manner which deviated from their perception of the standards of their subgroup made higher failure scores the more they were attracted to the group. Failure was defined in terms of the discrepancy between own performance and ideal performance. Festinger, Torrey, and Willerman (1954) varied the scores made by group members on an experimental task and obtained support for their hypothesis that the

> stronger the attraction of members to a group, the stronger will be the feelings of inadequacy on the part of those scoring less well than others and the stronger . . . [the] feeling of adequacy on the part of those scoring as well or better . . . [p. 173].

An experiment by Stotland and Hillmer (1962) has shown that identification is not necessarily related to attraction. Subjects who read that another person had been good at a clerical task evaluated their own performances on the same task more favorably than did subjects who read that the other person had done poorly. This effect was reported to be independent of the level of attraction to the model, although there was an overall preference for the "good" over the "poor" model on a "desire to converse with" question.

In general, the research on relationships between liking and self-evaluation has yielded equivocal results pointing to the need for further clarifying investigation.

Evaluation of the Situation

Marquis, Guetzkow, and Heyns (1951) have reported, from a field study of decision making conferences, significant positive correlations between a cohesiveness index and members' satisfaction with the group process and with the meeting. The cohesiveness index included observer ratings of liking among group members. Jordan (1953), in an investigation designed to test Heider's balance hypothesis, found that where two persons were described to subjects as liking one another, the situation which the hypothetical persons were in was rated as more pleasant than when the two persons were described as not liking each other. The relationship between persons had the most weight in influencing judgments of the situation, more weight than balance or imbalance in terms of agreement or disagreement between the two persons with regard to some object.

In a study by Exline (1957), subjects who had been told that they were in groups with well-matched, congenial others expressed greater satisfaction with their group's progress than subjects given the opposite orientation toward their fellow group members. Similarly, Libo (1953) had shown earlier that subjects told they were to work on a volunteered-for task with congenial group members differed significantly from subjects given the opposite instructions in their desire to stay with the group. Libo (1957) has also reported a significant relationship between the attraction of patients to their therapists, following their first interaction, and whether or not the patients returned for a second visit one week later.

That persons tend to respond favorably to situations in which liked others are present appears to be an empirically sound generalization.

Evaluation of Others

Perception of Similarity. A large number of investigations have provided evidence that persons will tend to perceive liked others as similar to themselves. The well known early study by Fiedler et al. (1952) confirmed the hypothesis that subjects will perceive fellow group members whom they like best as more similar to themselves (and their ideal selves) in personality than those whom they liked least. Corroborating data have subsequently come from samples of basketball and surveying teams (Fiedler, 1954), college freshmen (Kipnis, 1961), and sorority girls (Backman & Secord, 1962). Lundy, Katkovsky, Cromwell, and Shoemaker (1955) have also reported that personality descriptions of positive sociometric choices tend to be more similar to subjects' *acceptable* than to their unacceptable self-descriptions, while the reverse is the case for descriptions of negative sociometric choices. In a quite different research context, the investigation of marital conflict, happily married persons have been found to differ from unhappily married ones in attributing greater trait similarity to their spouses (Preston et al., 1952).

The same tendency to perceive preferred persons as being like oneself has been reported with respect to values and attitudes. Precker (1953) found that students attributed to their chosen advisers ratings of evaluative criteria which were similar to their own; and Davitz (1955) found that children at a summer camp tended to perceive their most preferred peers as liking the same camp activities that they did. Byrne and Wong (1962) obtained data indicating that highly prejudiced white subjects assumed greater attitude dissimilarity between themselves and an unknown Negro than between themselves and an unknown white person while white persons low in prejudice assumed the same degree of similarity with a Negro as with a white stranger. It has also been shown, among groups of Navy personnel, that persons tend to assume that individuals whom they like also like each other (Kogan & Tagiuri, 1958).

A number of other studies appear relevant here, although liking is not explicitly varied. Bieri (1953), for example, had some pairs of subjects discuss together and agree on conclusions about common experiences and preferred activities; control subjects wrote on these topics individually, back-to-back. Predictions about partners' responses to the Rosenzweig Picture Frustration Test were made before and after interaction. On the latter occasion, significantly greater similarity was found for the experimental subjects between their own responses and the responses which they predicted for their partners. In another investigation (Rosenbaum, 1959), subject pairs who had worked together cooperatively were reported to have subsequently assumed greater similarity with each other on personality traits than subjects who had worked competitively. Two studies (Smith, 1958; Stotland, Zander, & Natsoulas, 1961) have demonstrated that persons who are made aware of some similarity between themselves and another tend to perceive additional similarity.

Negative evidence has been provided by Lundy (1959). Subjects discussed a topic for five minutes, first with one student, then with another. During one of the interactions, subjects were led to believe that their partners had selected them for the discussion, while during the other, subjects were led to believe the opposite. Assumption of similarity with one's partner (on the Allport-Vernon value scales) was found to be a function of order of interaction (greater on the second interaction), but not to be associated with the acceptance-rejection variable.

Favorable Judgments of the Behavior of Liked Persons. Horowitz, Lyons, and Perlmutter (1951) reported that, in a face-to-face discussion group, agreement by group members with one another's acts was directly related to whether the "actors" were liked or disliked. This finding emerged only from questionnaire data, however, not from the verbal record obtained during actual discussion. In a different kind of situation, a summer camp for boys, it was found that the performance of high-status boys on an athletic task tended to be overestimated by their peers, while the performance of low-status boys was underestimated (Sherif et al., 1955). Variations in judgments of performance were not significantly related to actual skill displayed on the task.

Two studies by Berkowitz (1956, 1957) provide additional relevant data. High liking within effective air crews was found, in the first investigation, to be accompanied by members' perceptions of high motivation in one another. In the second investigation, utilizing simulated groups, high liking between partners (manipulated by means of instructions with regard to probable congeniality of one's fellow members) was accompanied by high regard for one another's opinions.

Perceptual Sensitivity. Suchman (1956) has reported, from a study of three five-person student groups, that "people who showed more favorability toward others were more accurate in estimating their feelings." Exline (1957) similarly found that group members who were told that they were well matched and congenial were more accurate than members of noncongenial groups in perceiving the task-oriented behavior of their fellow members. Also somewhat relevant is a study by Eisman and Levy (1960); they found that lip-reading was more accurate, the more the communicatee liked the communicator.

In a different kind of investigation Schutz (1961) composed groups so as to be homogeneous (compatible) with respect to members' interpersonal needs; after six meetings, group members were able to correctly identify the interpersonal characteristics of their group when the common characteristic of the members was expressed behavior rather than behavior desired from others.

From data obtained in studies not directly concerned with perceptual accuracy but with assumptions of similarity or judgments of others' behavior (cited earlier), it is evident that subjects sometimes assume too much similarity to persons they like or overestimate the goodness of their performance. Liking, therefore, may lead to inaccurate perception, although it may be associated with an increased sensitivity to the behavior and characteristics of liked others. Differential dependent measures of sensitivity and accuracy are needed.

Perceived Reciprocal Attraction. From a variety of groups in different research settings (Tagiuri, 1958; Tagiuri, Blake & Bruner, 1953; Tagiuri, Bruner, & Blake, 1958), Tagiuri and his co-workers have reported that the tendency toward "congruency," that is, the tendency to assume that one's positive or negative feelings toward another are reciprocated, is the most powerful relationship obtained and is significantly greater than would be expected on the basis of chance (a robot model). Tagiuri (1958) wrote:

> It is intriguing to ask whether congruency is a tendency for people to feel chosen by those whom they choose, or contrariwise, to choose those by whom they feel chosen. Undoubtedly, there are individual differences in the order of this cycle [p. 322].

In general, however, Tagiuri argued for the dominance of the former tendency, since individuals are usually first aware of their own feelings toward one another. Data relevant to the perception of "reciprocated attraction" have also been reported and discussed by Newcomb (1956, 1958, 1960).

Communication

The hypothesis that persons communicate more with liked than with disliked persons has been supported by a number of investigations but the results of other studies have indicated that in situations where opinions differ, and where there is pressure toward uniformity, individuals tend to increase their communication to deviates. In some of these latter investigations measures have indicated that the deviates are disliked or rejected while in others this must be assumed.

Kelley (1950) found that a larger proportion of students took part in a class discussion led by a substitute teacher when they had been informed that the teacher was "warm" than when they had been given the opposite information. Back (1951) compared high with low cohesive pairs (manipulated by instructions) with respect to communication patterns in discussion of a story derived from pictures and found significantly more "withdrawal" patterns among the latter. Grossack (1954) has reported that more notes were exchanged by subjects under instructions to "cooperate" with fellow group members than by subjects instructed to "compete" on an experimental task. In still another relevant experiment by Cervin (1956), the "solidarity" of three person groups was manipulated by having stooges agree or disagree with opinions expressed by the critical subject. Dependent measures indicated that both average degree of participation in discussion and response latency of critical subjects was greater in the "solidarity situation." Utilizing natural student groups, Lott and Lott (1961) obtained a significant positive correlation between cohesiveness, as measured by an index of interpersonal attraction, and median frequency of communication among group members on a discussion topic supplied by the investigators. Finally, Runkel (1962) has reported finding a relationship between the level of respect existing between secondary school teachers and how often they communicate with one another; the causal direction of this relationship, however, is not clear.

An investigation by Mellinger (1956) has provided interesting data not on frequency of communication but on the nature of communication as a function of interpersonal attraction. It was found that scientists in a research organization tended to conceal their attitudes about a particular issue when communicating with persons in whom they lacked trust.

In studies of small groups under pressure to achieve uniformity on a given question or problem, it has generally been found that communication tends to be directed more to deviants than to other group members, especially in relatively cohesive or homogeneous groups (Emerson, 1954; Festinger et al., 1952; Festinger & Thibaut, 1951; Schachter, 1951). More recently, Altman and McGinnies (1960) have reported that both heightened pace of discussion and frequency of opposition-directed remarks were positively related to members' valuations of their groups.

In the investigations cited in the foregoing paragraph, communication directed by group members to deviants consisted, primarily, of attempts to influence a change of opinion. It is probable, then, that the relationship between interpersonal liking and communication will vary with the conditions under which the interaction takes place.

Uniformity-Conformity

Most of the findings support a positive relationship between the degree of interpersonal attraction existing among persons and their consequent uniformity with regard to particular opinions, attitudes, judgments, or other behavior. Both attempts *to* influence and successful influence *by*, liked persons have been shown. These two processes, which appear logically and empirically to operate together, are usually not separated, but some investigators have obtained measures of influence attempts as well as of resultant changes in behavior.

A study by Back (1951) is often cited in support of the hypothesis that pressures toward uniformity are greater in high than in low cohesive groups although Back's data provide only minimal support. Pairs of individuals worked on a story-writing task under different instructions with regard to personal attraction, task direction, or group prestige, but whether the attempted variations in cohesiveness were successful is not clear. When both partners were taken into account, no significant differences were obtained between the high and low cohesive pairs on a measure of change following influence attempts, although in high cohesive groups there was reliably more change on the part of one partner.

The technique of manipulating group cohesiveness by telling subjects that they will probably like (find congenial) or not like one another has been used in a large number of investigations of the attraction-uniformity hypothesis. Festinger et al. (1952) told members of some groups that they would find each other compatible, congenial, and interesting, and would most likely get along very well. Such groups were found to exert greater pressures for uniformity of member opinion with regard to a particular issue under discussion than groups given the opposite orientation. Similar instructions were used by Berkowitz (1957) in a study where subjects worked under simulated group conditions, and conformity to the responses of partners on a task was found to be significantly related to liking of the partners. In an earlier investigation Berkowitz (1954) had related intermember liking with the persistence of group productivity standards for an experimental task similar to one utilized by Schachter, Ellertson, McBride, and Gregory (1951). This latter study found that persons in high cohesive groups followed presumed member suggestions for low productivity to a significantly greater extent than did members of low cohesive groups, but when the pressure was for high production, cohesiveness made no difference.

Other research in which interpersonal liking has not been explicitly manipulated may be considered relevant to the present discussion if we assume variation in the level of attraction among subjects. Thus, for example, Bovard (1951b) found significantly more convergence in perceptual judgments among members of group-centered classes than among members of leader-centered ones. These two types of classes differed in amount of student interaction. Similarly, Festinger and Thibaut (1951) reported that among group members told that they were with others similar to themselves in interest in and knowledge about a given problem there was greater change in the direction of opinion uniformity than among group members given the opposite instructions. Homogeneity was associated with increased uniformity only for opinions on a football strategy problem, however, and not for those concerning treatment of a delinquent boy. In a study by Grossack (1954), some subjects were instructed to work competitively and others cooperatively. The findings provided support for the experimental hypothesis that cooperating individuals (assumed by Grossack to manifest a cohesive group structure) will make more attempts to influence others and will accept more pressures in the direction of uniformity than competing persons. Kidd and Campbell (1955) varied the success of groups on an anagrams task and found, subsequently, that subjects in "success" groups differed significantly from those in "failure" groups in conforming more to purported group averages regarding estimated light flickers. Finally, Dittes and Kelley (1956) have reported that group members who received ratings of average acceptability from their fellows tended to show more conformity behavior than those who received low or very high ratings.

Another group of studies have provided data on differences in conformity behavior between groups which more clearly differed in level of interpersonal attraction. Thrasher (1954) investigated variations in perceptual judgment as a function both of gradations in stimulus structure and interpersonal relations. He found that the judgments of pairs of friends "were significantly more influenced by each other's judgments than the judgments of pairs of individuals who were not so positively involved (neutrals)." This was the case for stimuli of both intermediate and low structure value. From a quite different research setting, a field study of secondary school teachers, Rasmussen and Zander (1954) have reported that the discrepancy between perceived group standards (regarding teaching performance) and own ideal was smaller, the more attracted was a teacher to his group. From another field study, of decision-making conferences, it was concluded that the achievement of high consensus is associated with warm and friendly personal feelings among the participants (Guetzkow & Gyr, 1954). Thibaut and Strickland (1956) compared conformity behavior within groups of positively associated individuals (fraternity pledges) and within groups of strangers. Under varying pressure of believing oneself to be in the minority with regard to judgments, the greatest positive relationship between pressure and conformity was found to occur in those "pledge" groups which had been given a "group" as opposed to a "task" orientation. A more recent investigation by Zander, Natsoulas, and Thomas (1960) reported that subjects who were most attracted to their group were also ones whose own aspiration level (or personal goal), with regard to an experimental task, was highly congruent with the goal of their group. In an investigation of natural student groups, Lott and Lott (1961) obtained a significant relationship between the degree of mutual positive attitudes existing among group members and their conformity to a purported group standard with respect to an opinion question.

One experimental approach to the question of uniformity pressures and group cohesiveness or of intermember attraction, has been the demonstration of less tolerance for, or greater rejection of, deviates in high cohesive groups. Schachter (1951) obtained data supporting this hypothesis in a study where cohesiveness was manipulated through the use of differential instructions and stooges expressed points of view at variance with majority opinion. Generally positive but not definitive findings were also obtained from replications of this study by Emerson (1954) and by Schachter, Nuttin, de Monchaux, Maucorps, Osmer, Duijker, Rommetveit, and Israel (1954). Similarly, Festinger et al. (1950), who studied two parts of a student housing project, reported that in one there was a high, significant, negative correlation between a sociometric measure of cohesiveness and the percentage of deviates (with respect to a particular issue) in residence courts, and that deviates in these courts tended to receive fewer sociometric choices than those who shared the majority opinion. In the other portion of the project, however, neither of these findings was obtained.

Still another empirical variation has been the demonstration that individuals are more resistant to changing their opinions if these are shared by members of groups to which they are attracted. Gerard (1954), for example, reported that a paid participant who challenged the opinions of subjects, each of whom represented a group, was more effective in influencing those who were less attracted to their fellow group members. In an earlier investigation, Kelley and Volkart (1952) had obtained similar data from a natural setting. A sample of Boy Scouts heard a speaker criticize the Scouts' emphasis on camping and woodcraft. In general, those boys who valued their Scout troop membership least changed most in the direction of negative attitudes toward these activities.

We have classed together a group of studies in which influencer (model) and influencee (imitator) are distinct and clearly identifiable. Taken together, the results of these investigations provide considerable support for the proposition that an individual will tend to match his behavior to that of a liked other. In a laboratory study, Grosser, Polansky, and Lippitt (1951) had pairs of boys work together on block-designs. One of the boys, an experimental collaborator, varied his behavior with different partners so as to produce the conditions of high and low friendliness. Under the former condition as compared with the latter, there was more imitation by the subjects of certain aspects of the collaborator's behavior during rest periods. Lippitt, Polansky, and Rosen (1952) have also reported, from a field study at a children's camp which replicated earlier research, that the boys investigated were more likely to imitate the behavior of those to whom they attributed high power and that attributed power choices were highly related to judgments of personal liking. Harvey and Consalvi (1960) had groups of boys from a school for delinquents participate in a perceptual judgment task and found, in general, that the higher a boy's status (i.e., the more preferred he was as a teammate or team captain), "the greater his proportionate influence on members' judgments . . ." Another investigation of children at a summer camp (Cohn, Yee & Brown, 1961) obtained measures of food and activity preferences and of interpersonal attraction. Over a 4-week period, a significant change in preferences was found in the direction of similarity with most liked peers. In addition, a high degree of correspondence was noted between actual eating behavior at meal time and measured attitude changes. Change in one's self-evaluation so as to match the characteristics perceived in one's best friend has also been reported, among a sample of male college freshmen (Kipnis, 1961).

Some studies have shown imitation of a stranger, or presumed previously neutral individual, under conditions designed to increase his attractiveness. For example, in one investigation (Kelley & Woodruff, 1956) college students listened to a recorded speech which advocated opinions contrary to the norms of their college. Some were told that the audience heard applauding the speaker consisted of faculty and alumni from the college; others were informed that the applauding audience was made up of anonymous outsiders. More opinion change in the direction advocated by the speaker took place under the former than under the latter condition. Burdick and Burnes (1958) manipulated another variable and found that a lecturer, after speaking pleasantly for an hour to a group of students, was effective in changing their opinions, while after speaking unpleasantly on a later occasion, he was significantly less liked than previously and students' opinions tended to change in a direction opposite to his. (In the same paper Burdick and Burnes reported finding a positive relationship between emotionality, as measured by the galvanic skin response, and disagreement with a liked experimenter.) Burnstein, Stotland, and Zander (1961) have reported that grade-school children tended to model their preferences regarding deep sea diving more after an adult described as highly similar to them in background and other attributes than after a less similar adult. Under conditions where subjects were not informed about the adult's degree of similarity to themselves, but were told about his sea diving ability, preferences of the model were more accepted by the children as their own when he was described as capable than when he was described otherwise. Suggestive data have been reported by McDavid (1959) from a study of imitation among preschool children: A female model was found to be initially more effective in providing behavioral cues than was a male model. Another and more recent investigation of nursery school children (Bandura & Huston, 1961) varied the behavior of the female experimenter so that some of the children experienced two warm and nurturant play sessions, while the others experienced two cold, nonnurturant ones. There was more imitation of the experimenter's verbal and motor behavior by children in the former condition (except for aggressive behavior which was readily imitated by all subjects).

Although the empirical evidence favoring a positive relationship between interpersonal attraction and consequent behavior similarity is extremely impressive, there have been negative findings as well. Bovard, who had reported a positive relationship between interaction and convergence on perceptual judgments in one study (1951b), found in a later one (1953) that conformity on another perceptual judgment task was not reliably related to a measure of the average attraction of group members to their group. McKeachie (1954) compared group-centered with leader-centered classes and obtained equivocal results with respect to congruence between individual attitudes and perceived and actual group norms. Downing (1958) has reported failure to obtain a significant difference in level of conformity between members of high and low cohesive groups in an autokinetic response situation.

In a comparison among industrial work groups differing in cohesiveness, Seashore (1954) found that high and low groups did not differ reliably in within group variance with respect to members' judgments about a reasonable production level, but the within group variance in high cohesive groups was significantly less than in low groups with respect to *actual* production. Burdick (cited by Walker & Heyns, 1962, pp. 30–40), has also reported both negative and positive findings under different conditions. High liking for other group members was found to be positively related to *nonconforming* behavior in one experiment and to *conforming* behavior in a second experiment. In the latter case,

> conformity was made instrumental to being liked by other persons . . . by adding to the instructions the point that it is common knowledge that persons tend to like one another more, to get along better, when they agree about common problems [p. 34].

Task Performance

As was the case with respect to uniformity of behavior, there exists an extensive literature devoted to the relationship between interpersonal attraction and productivity.

One experimental technique has simply compared groups of friends (or positive sociometric choices) with groups of nonfriends on some measure of productivity. An early study by Husband (1940) found superior performance by pairs of sixth- and seventh-grade friends on word puzzles, jigsaw puzzles, and arithmetic problems. At three summer camps in which workgroups were formed according to choices made so that congeniality was maximized, "Observers reported a superior solidarity over former camps as reflected in such indices as greater participation in discussions . . . and more work accomplished [Faunce & Beegle, 1948]." Van Zelst (1952a, 1952b), experimenting in a natural work situation, grouped some carpenters and bricklayers into mutual choice pairs and found that they excelled over control pairs on such indices as savings in labor costs and materials, rate of turnover, and job satisfaction. A more recent laboratory investigation (Bjerstedt, 1961) found that groups of children composed of chosen workmates surpassed groups composed of nonchosen children on story-telling tasks; they made higher scores on indices of interest in the quality of performance (process time and amount of rehearsal), and they achieved better results.

Other studies have obtained correlations between measures of interpersonal attraction and performance on some particular task. Goodacre (1951, 1953), for example, reported significant positive correlations between various indices of group cohesiveness and the scores made by combat and rifle squads on field problems. Similarly, Berkowitz (1956) has reported a significant relationship between aircrew members' attitudes toward one another and two criteria of their actual combat effectiveness; and Chapman and Campbell (1957), in a laboratory study, found that, "Ratings of a person's desirability as a teammate correlate significantly with the success of the teams of which he is a member." Positive findings have also been reported by Darley, Gross, and Martin (1952) from a study of women's residence units and by Gardner and Thompson (1956) from a study of fraternities.

Supportive findings may also be cited from investigations in which the independent variable may be assumed to reflect variation in interpersonal attraction; in some cases such variation has been shown, explicitly, to accompany the manipulation of conditions. For example, Sacks (1952) administered individual intelligence tests to three groups of nursery school children. The tester had previously established a "good" relationship with the children in one group and a "poor" relationship with the children in a second group, and she had had no prior contact at all with the remaining children. The scores of the children who were familiar with the tester were significantly higher than they had been 10 days earlier on an alternate form of the test whereas this was not the case for the control children; moreover, the retest increases of the "good" relationship children were significantly greater than the increases of the "bad" relationship children. According to Sacks, qualitative observational data support the fact that the children differed in the intended way in their attitude toward the experimenter. In a study of a chain of supermarkets, Clark (cited by Homans, 1961, p. 255) obtained an almost perfect correlation between an index of social ease (status congruency of "bundler" and "ringer") and an index of labor efficiency. Cervin (1956) varied group "solidarity" by having two stooges reinforce or extinguish the opinions of the critical member of each group. The same subjects "showed better and more even performance" when in the solidarity than in the dissolidarity condition. The compatibility of groups, defined by similar interpersonal orientations among group members, was varied by Schutz (1955) and by Moos and Speisman (1962). Data from both studies show a tendency toward better task performance by compatible groups. Schutz's results also indicated that the difference between compatible and incompatible groups varied with the type of task. In an investigation by E. E. Smith (1957), group productivity was shown to be negatively affected by the presence of two silent members in five-person experimental groups; the silent members were perceived as being unpredictable.

That individuals who work together cooperatively are more productive on some criterion than group members who compete against one another has been demonstrated for classroom groups by Deutsch (1960) and by Smith, Madden, and Sobol (1957). Blau (1954), interested in the same variable, compared groups of employment agency interviewers and found, paradoxically, that competitiveness and productivity were "inversely related for groups but directly related for individuals in the competitive group." The competitive group was less productive than the group in which there were cooperative norms and in which competitive tendencies were curbed, but within the former group, the more competitive individuals were relatively more productive.

Not directly concerned with actual performance, but with individual aspiration levels for group achievement, is an investigation by Ex (1959). Subjects who worked with a sympathetic partner (a congenial confederate) were found to have higher levels of aspiration for their joint performance than subjects who worked with an unsympathetic partner.

Not all of the research findings in this area have pointed to the same conclusion. In an early study by Philp (1940), for example, in which pairs of kindergarten children who preferred each other or were strangers worked on simple tasks, the "quality of the response made during the two types of pairings" was reported to be very different, but there were no reliable differences in average performance or range of performance. In a far different setting, a field study of actual conferences, cohesive groups were found to be no more productive than less cohesive ones (Marquis et al., 1951), and among industrial work groups a tendency was found for either high or low production among high cohesive groups, depending upon the group's standard (Seashore, 1954). That high cohesive groups can be induced to lower production or perform poorly has been demonstrated in experiments by Schachter et al. (1951) and by Berkowitz (1954).

Atthowe (1961), who measured the cohesiveness of decision-making pairs by their number of "we" versus "I" remarks, also found no relationship between this measure and task performance, while Adams (1953) has reported a curvilinear relationship between the degree of status congruence within bomber crews (which varied positively with friendship) and measures of their technical performance. It has also been suggested that under certain conditions there is a negative relationship between interpersonal liking and efficient task performance. According to Fiedler (1953), "quite in contrast to therapeutic and social relations, we find that psychological closeness and warmth in key members of small groups seems to be a detriment to effective team work . . ." under conditions where teams "want to get a job done." Similarly, Stogdill (1959) has suggested that "the effort that is devoted to the development of integration might be conceived as a subtraction from the efforts that are devoted to productivity [p. 269]." Supporting data may be found in investigations of machine operators in a shoe factory (Horsfall & Arensberg, 1949), of surveyor teams (Fiedler, 1954), and of rifle teams (McGrath, 1962). Hoffman (1959) and Hoffman and Maier (1961) have varied the homogeneity of members' personalities in problem-solving groups and found that nonhomogeneous groups tended to produce superior and more inventive solutions. In these studies data on intermember liking were not obtained.

Equivocal results have been reported by Yuker (1955), who used group recall as a measure of group productivity. In an interesting experiment, the leadership (democratic versus authoritarian) and cooperation-competition dimensions of groups were varied, and individuals recalled alone and collectively a story which had been read to them while they were assembled in their four-person groups. The difference in accuracy between group recall and initial individual recall was found to be significantly greater in cooperative than in competitive groups, but there was no difference between democratic and authoritarian groups, and neither the cooperation-competition nor the leadership variable was found to affect differences in accuracy between final individual recall and initial individual recall. Yuker's data, however, suggest that his groups did not differ reliably in cohesiveness. In a subsidiary analysis, Yuker found that the 10 groups in which there had been the greatest improvement from initial to final recall differed from the 10 groups in which there been the least improvement on three questionnaire items reflecting cohesiveness.

Findings relevant to the relationship between liking and task performance are contradictory. It seems likely that in a task situation other variables such as the demands of the situation itself (instructions or job specifications), the standards of performance preferred by liked co-workers, and the degree to which sociability may interfere with the required behavior for a particular job, may be highly significant.

Learning

We are not concerned here with the question of whether learning in a group situation is superior to learning when alone, a question discussed often in the earlier literature (see Lorge, Fox, Davitz, & Brenner, 1958) and more recently by Gurnee (1962). Our interest is rather in those investigations which compare groups differing in the level of attraction among members or in some related variable. Such studies have been conducted in a variety of learning situations and have yielded generally equivocal results.

The major independent variable in studies by Perkins (1950) and Rehage (1951) was group- versus leader-centered classroom climates. In the former, positive results were obtained with respect to the learning of child development material by in service teachers; in the latter, eighth-grade students under the "teacher-pupil planning" condition were found not to differ significantly from those under the "teacher planning" condition in the "amount of social studies subject matter learned." Calvin, Hoffman, and Harden (1957) investigated the effect of climate (permissive, democratic versus nonpermissive, authoritarian) on group problem solving. The trends in three experiments supported their hypothesis that a permissive climate is superior to a traditional one for subjects with high intelligence, but a handicap for subjects with only average intelligence.

In a recent investigation by Shaw and Shaw (1962) the sociometric composition of three-member groups of second-grade children was varied. The children, in their groups, studied lists of spelling words and were tested during both an early and late period of interaction. Spelling scores were found to be positively correlated with an index of group cohesiveness or intermember liking only on the first test; they were unrelated on the second test.

A number of relevant studies have been concerned with verbal conditioning as a function of the relationship between experimenter and subject, or simply the attractiveness of the experimenter. Binder, McConnell, and Sjoholm (1957) found a higher rate of learning (for "mildly hostile" words) by subjects whose reinforcements were provided by an attractive, soft-spoken young lady than by subjects reinforced by a very masculine, former Marine captain. No data were provided on the subjects' attitude toward the two experimenters. On the other hand, negative findings have been reported by Kanfer and Karas (1959). Prior to the conditioning of "we" or "I" some subjects had been tested by the experimenter and told that they did well, or poorly, or told nothing, while some subjects were given no prior experience with the experimenter. Those subjects who had interacted with the experimenter conditioned significantly better than those in the no experience group but there were no reliable differences among the former in rate of conditioning or final response level. Although questionnaire results indicated that the "failure" subjects considered the experimenter to be unfriendly and a source of discomfort, these subjects also said that they were more highly motivated on the conditioning task. Using a similar design, Simkins (1961) obtained similar results. Hostile words were reinforced after subjects had had prior experience with the experimenter. When the reinforcement consisted of the experimenter's saying "good" or "that's fine," the "failure" subjects (i.e., those previously criticized by the experimenter) conditioned better than any other group of subjects; when points or pennies were used as reinforcers, conditioning was not related to the experimenter's prior behavior.

Weiss, Krasner, and Ullman (1960) also varied the experimenter's behavior to induce a hostile or a positive atmosphere prior to a critical series of conditioning trials. The hostile condition served to decrease responsiveness but did not affect reconditioning. Quite similar findings were reported by Reece and Whitman (1962). In this study the experimenter varied his expressive movements *during* the conditioning procedure and successfully induced perceptions of "warmth" and "coldness" by different subjects. Those under the "warmth" condition produced significantly more words, but the number of plural nouns (which were reinforced) was not affected.

Still another kind of manipulation was employed by Sapolsky (1960). Some subjects were told that they would be working with a congenial experimenter, while others were given the opposite instructions. A measure of the success of the manipulation yielded positive findings, and response acquisition was found to be significantly greater for high- than for low-attraction subjects. In a second study, Sapolsky varied the "compatibility" of subject-experimenter pairs by selecting them on the basis of scores on Schutz's Fundamental Interpersonal Relations Orientation-Behavior questionnaire and again obtained positive results. These data provide the clearest support for the proposition under consideration.

We must cite, finally, some unique findings reported by Horwitz (1958) who has conducted a series of investigations on the effects of the *veridicality* of one's expressions of like or dislike. In one study, a teacher deferred to students' desires under one condition and was, therefore, power *enhancing* with respect to the students, while in another condition the teacher was power *reducing* and always acted upon his own desires. These manipulations, designed to produce liking or disliking of the teacher, had proved successful in previous investigations. Then, by manipulating information about group attitudes toward the teacher, either a veridical or nonveridical group norm was developed. Subsequently, on a type-setting task, the number of trials required to achieve two consecutive perfect performances was found to be significantly lower (i.e., faster learning occurred) under the veridical than the nonveridical condition. In other words, whether the students liked or disliked the teacher did not matter so long as the group norm was the same as the attitudes which were generated by the experimental manipulation of the teacher's behavior.

Overview

Compared with the empirical findings regarding the antecedents of liking, those relevant to consequences are far less clear-cut and definitive. This is reflected in the relative lack of theoretical concern with the effects of attraction, as we shall see below. Since applications of so-called group behavior principles are often urged in such fields as group psychotherapy, education, and community relations, it is vital to distinguish between validated and unvalidated hypotheses or to differentiate between principles which have been relatively well substantiated by research and those for which the empirical evidence is weak or contradictory. In the former category might be placed generalizations concerning the effect of interpersonal attraction on evaluation of the situation, on evaluation or perception of persons who are liked, and on uniformity of behavior or opinions on *certain* issues, perhaps those that are most relevant to the personal relationship. Hypotheses linking interpersonal liking with the expression of aggression, self-evaluation, communication, task performance and learning require additional refinement. In need of systematic coordination within some general framework are all of the variables which we have considered here.

CONSEQUENCES OF LIKING: THEORETICAL EXPECTATIONS

It is surprising to find that so little has been written of a systematic nature regarding the predictable effects of liking between individuals or among members of a group. Furthermore, in much of what has been written it is difficult to identify underlying general principles; the predictions seem, for the most part, to be more in the form of empirical generalizations, and consequently, one theoretical position cannot easily be distinguished from another. There are exceptions and some attempts at genuine derivation from sets of limited, broader principles.

Festinger

Festinger has not been concerned with interpersonal attraction, per se, but with a number of related social phenomena. He predicted, for example (Festinger, 1950a), that as forces to remain within a group increase (i.e., as the group's attractiveness increases), pressures to communicate will also increase. Social communication, according to Festinger, is a means by which individuals compare their opinions with one another (thereby validating them against the standard of social reality) and is a means by which individuals may more effectively progress toward a group goal. Both of these communication-provoking needs (or forces) are presumably greater in high-cohesive or more attractive groups. Elsewhere, Festinger (1954) more explicitly hypothesized that the greater the attractiveness of a group, the more important it is as a standard of comparison with respect to abilities and opinions. Further, the amount of opinion change on the part of an individual is a function of the forces acting on him to remain in the group, or, in other words, is a function of the attractiveness of the group, its members, or activities.

Heider

On the basis of the balance concept which underlies his predictions of interpersonal behavior, Heider (1958) has proposed that when one person likes another the following consequences will be probable: (*a*) Similarity will increase between the two persons with respect to their attitudes toward objects, events, or other persons, resulting from changes in the liker, changes in his perception of the object of his liking, or both; (*b*) The liked person will be approached or contacted and the liked person will be benefited or helped and will be perceived as benefiting the liker.

Newcomb

In Newcomb's theoretical papers (especially 1953) can be found considerable explicit discussion of the effects of interpersonal liking. His "strain toward symmetry" proposition leads to the prediction that the greater the valence of one person (B) for another (A), or "the more intense one person's concern for another," the greater is the likelihood that A will be sensitive to B's "orientations to objects in the environment," and that symmetry of orientations will be achieved. Perceived symmetry, as well as the achievement of objective symmetry, will vary as a function both of "intensity of attitude toward [issue] X and of attraction toward B."

An important refinement in Newcomb's position is represented by his hypothesis that there will be a greater range of subjects or issues "with regard to which there is strain toward symmetry" between persons more attracted to one another than between persons less so, and for the latter "strain toward symmetry is limited to those particular Xs, co-orientation toward which is required by the conditions of association." In a later paper, Newcomb (1960) proposed, in addition, that A will tend to endow B with "favorable attributes" if A is positively attracted or has a favorable attitude toward B.

Thibaut and Kelley

These theorists (Thibaut & Kelley, 1959) are concerned primarily with the consequences of one person's *dependence* upon another, rather than with one person's attraction to another. Although dependence and attraction are seen as tending to be positively correlated, the concepts are independently defined and related to other variables. High cohesiveness, for example, is said to result from a condition of high interdependence among group members, a condition in which each member has high power over the other, or

> can move the other through a relatively great number of reward-cost units above [p. 114]

the other's comparison level based on alternative group relationships. It follows, then, that in

> high cohesiveness both the ability to make demands and the ability to resist them are greater than in low cohesiveness [p. 114],

producing a potentiality for conflict under the former condition. Cohesive groups, however, quickly find a procedure to avoid conflict: one member will have more power than others, or the members define a range of issues on which conformity is expected. When members have power over one another, this means that they have the ability to reward each other; this ability

> seems to create further ability to do so . . . A will reward B in order to get B to produce behavior that he, A, finds rewarding. By so doing, A creates conditions under which, through association learning, B is likely to learn to like this behavior himself. The greater A's ability to reward B, the more this learning will occur [p. 115].

In this way, members of cohesive groups move toward similarity in values, attitudes, and behavior.

Other Approaches

With positive attitudes among persons defined in learning-theory terms as anticipatory goal responses having cue and drive properties, Lott (1961) has predicted consequences for group behavior derived from the Hullian framework. Since attitudes have drive value, the greater the degree of intermember liking the higher should be the general level of group activity, the quantity (not necessarily the quality) of task production, and the level of communication. Similarly based on the assumed drive value of attitudes is the prediction that more efficient learning should take place in high than in low cohesive groups or on the part of individuals who are in the presence of liked others.

It is also predicted that the "more cohesive the group the greater the probability that members will develop uniform opinions and other behaviors with respect to matters of consequence to the group." This hypothesis stems from the fact that stimuli which evoke anticipatory responses are thus able to function as secondary reinforcers since this property of a stimulus depends upon the stimulus being able to evoke an anticipation of reward. Persons who are liked, in other words, are sources of secondary reward. It can, therefore, be expected that the greater the interpersonal attraction among persons, the greater is their reinforcing potential, and the greater will be "the probability of uniformity among them on issues where deviancy is accompanied by withdrawal of group members from the deviant." Thus, uniformity is not always expected with respect to all issues. Lott and Lott (1961) have suggested further that a positive relationship between cohesiveness and conformity can be predicted

> only in those situations where the stimulus features are similar, to some degree, to those which were present when group members acquired their secondary reinforcing characteristics. The principle of stimulus generalization, in other words, should be employed to predict the effectiveness with which group members can act as secondary reinforcers for one another in diverse situations [p. 411].

Some General Considerations

A number of years ago Albert (1953) raised provocative questions regarding the concept of group cohesiveness. He discussed the generality of its conceptual definition and the diverse and questionable attempts to operationalize it and asked, "is cohesiveness necessary?" We believe that at the present stage of development within the area of group behavior, we would do better to investigate more clearly definable factors such as interpersonal attraction, establishing relationships with antecedent and consequent variables. We have attempted here a beginning in this direction. Cartwright and Zander (1960) have suggested a similar approach:

> The most important task for group dynamics, is to establish a generally accepted set of basic variables and concepts having clear empirical and conceptual meaning. . . . One part of the problem is to iso-

> late the actual unitary variables or dimensions that make discernible differences . . . The differentiation of unitary variables can only be accomplished by empirical work which discovers what regularities are invariably found among measurements and observations [p. 46 f.].

We do not intend to throw out group concepts and to substitute for them concepts definable in terms of the behavior of individuals, but we believe it to be both more parsimonious and systematic to *begin* with the latter, our only source of data, after all. Motivation for group membership resides in the individual, and the consequences of group processes are seen in changes in individual behavior. Our thinking although primarily influenced by modern learning theory, is in accord with other contemporary positions within social psychology. Homans (1961), for example, maintains that "from the laws of individual behavior . . . follow the laws of social behavior when the complications of mutual reinforcement are taken into account [p. 30]." Floyd Allport's (1962) theoretical orientation is similar:

> When the group dynamicist speaks of the "attraction of the group for the individual" does he not mean just the attraction of the individuals for one another? If individuals are all drawn toward one another, are they not *ipso facto* drawn to the group [Pp. 23–24]?

Allport has refined his early theoretical position which saw the explanation of social phenomena *"only in the psychology of the individual as he operated in situations with others* [p. 5]" to a view of groups as collective "structures" of articulated behaviors engaged in by interacting persons. In other words, what constitutes a group are *"particular give-and-take behaviors* [p. 13]."

A decade has passed since Riecken and Homans' (1954) excellent review of literature relevant to the "psychological aspects of social structure," but the need for codification of research findings, which they emphasized, is still a pressing one. Our aim, too, has been to place investigations under a "rationally ordered set of headings." We also share with Riecken and Homans the assumption that "there are no such things as contradictory findings," that such findings must have been reached in different circumstances and that the search to identify the specific characteristics of these circumstances "is one of the most potent sources of growth in knowledge."

REFERENCES

Adams, S. Status congruency as a variable in small group performance. *Social Forces,* 1953, **32**, 16–22.

Albert, R. S. Comments on the scientific function of the concept of cohesiveness. *American Journal of Sociology,* 1953, **59**, 231–234.

Allport, F. H. A structuronomic conception of behavior: Individual and collective. I. Structural theory and the master problem of social psychology. *Journal of Abnormal and Social Psychology,* 1962, **64**, 3–30.

Altman, I., & McGinnies, E. Interpersonal perception and communication in discussion groups of varied attitudinal composition. *Journal of Abnormal and Social Psychology,* 1960, **60**, 390–395.

Aronson, E., & Mills, J. The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology,* 1959, **59**, 177–181.

Atthowe, J. M. Interpersonal decision making: The resolution of a dyadic conflict. *Journal of Abnormal and Social Psychology,* 1961, **62**, 114–119.

Back, K. Influence through social communication. *Journal of Abnormal and Social Psychology,* 1951, **46**, 9–23.

Backman, C., & Secord, P. The effect of perceived liking on interpersonal attraction. *Human Relations,* 1959, **12**, 379–384.

Backman, C., & Secord, P. Liking, selective interaction, and misperception in congruent interpersonal relations. *Sociometry,* 1962, **25**, 321–335.

Bandura, A., & Huston, Aletha C. Identification as a process of incidental learning. *Journal of Abnormal and Social Psychology,* 1961, **63**, 311–318.

Bass, B. Increased attraction to the group as a function of individual and group goal attainment. In, Behavior in groups. Technical Report No. 2, May 1955. (Mimeo)

Bass, B. Some observations about a general theory of leadership and interpersonal behavior. In L. Petrullo & B. Bass (Eds.), *Leadership and interpersonal behavior.* New York: Holt, Rinehart & Winston, 1961. Pp. 3–9.

Berkowitz, L. Sharing leadership in small, decision-making groups. *Journal of Abnormal and Social Psychology,* 1953, **48**, 231–238.

Berkowitz, L. Group standards, cohesiveness, and productivity. *Human Relations,* 1954, **7**, 509–519.

Berkowitz, L. Group norms among bomber crews: Patterns of perceived crew attitudes, "actual" crew attitudes, and crew liking related to aircrew effectiveness in Far Eastern combat. *Sociometry,* 1956, **19**, 141–153.

Berkowitz, L. Liking for the group and the perceived merit of the group's behavior. *Journal of Abnormal and Social Psychology,* 1957, **54**, 353–357.

Berkowitz, L. The expression and reduction of hostility. *Psychological Bulletin,* 1958, **55**, 257–283.

Berkowitz, L., & Green, J. The stimulus qualities of the scapegoat. *Journal of Abnormal and Social Psychology,* 1962, **64**, 293–301.

Berkowitz, L., & Holmes, D. S. A further investigation of hostility generalization to disliked objects. *Journal of Personality,* 1960, **28**, 427–442.

Berkowitz, L., Levy, B., & Harvey, A. R. Effects of performance evaluations on group integration and motivation. *Human Relations,* 1957, **10**, 195–208.

Berkowitz, L., & Macauley, Jacqueline. Some effects of differences in status level and status stability. *Human Relations,* 1961, **14**, 135–148.

Bieri, J. Changes in interpersonal perceptions following social interaction. *Journal of Abnormal and Social Psychology,* 1953, **48**, 61–66.

Binder, A., McConnell, D., & Sjoholm, Nancy. Verbal conditioning as a function of experimenter characteristics. *Journal of Abnormal and Social Psychology,* 1957, **55**, 309–314.

Bjerstedt, Ake. Preparation, process, and product in small group interaction. *Human Relations,* 1961, **14**, 183–189.

Blau, P. Cooperation and competition in a bureaucracy. *American Journal of Sociology,* 1954, **59**, 530–535.

Bonner, H. *Group dynamics: Principles and applications.* New York: Ronald Press, 1959.

Bovard, E. W. The experimental production of interpersonal affect. *Journal of Abnormal and Social Psychology,* 1951, **46**, 521–528. (a)

Bovard, E. W. Group structure and perception. *Journal of Abnormal and Social Psychology,* 1951, **46**, 398–405. (b)

Bovard, E. W. Conformity to social norms and attraction to the group. *Science,* 1953, **118**, 598–599.

Bovard, E. W. Grouping error and interpersonal affect: A correction. *Journal of Abnormal and Social Psychology,* 1956, **52**, 283–284. (a)

Bovard, E. W. Interaction and attraction to the group. *Human Relations,* 1956, **9**, 481–489. (b)

Bowerman, C. E., & Day, Barbara. A test of the theory of complementary needs as applied to couples during courtship. *American Sociological Review,* 1956, **21**, 602–605.

Broxton, June. A method of predicting roommate compatibility for college freshmen. *Journal of the National Association of Women Deans and Counselors,* 1962, **25**, 169–174.

Bugental, Daphne, & Lehner, G. F. Accuracy of self-perception and group perception as related to two leadership roles. *Journal of Abnormal and Social Psychology,* 1958, **56**, 396–398.

Burdick, H. A., & Burnes, A. J. A test of "strain toward symmetry" theories. *Journal of Abnormal and Social Psychology,* 1958, **57**, 367–370.

Burnstein, E., & McRae, Adie. Some effects of shared threat and prejudice in racially mixed groups. *Journal of Abnormal and Social Psychology,* 1962, **64**, 257–263.

Burnstein, E., Stotland, E., & Zander, A. Similarity to a model and self-evaluation. *Journal of Abnormal and Social Psychology,* 1961, **62**, 257–264.

Byrne, D. The influence of propinquity and opportunities for interaction on classroom relationships. *Human Relations,* 1961, **14**, 63–70. (a)

Byrne, D. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology,* 1961, **62**, 713–715. (b)

Byrne, D. Interpersonal attraction as a function of affiliation need and attitude similarity. *Human Relations,* 1961, **14**, 283–289. (c)

Byrne, D., & Buehler, J. A. A note on the influence of propinquity upon acquaintanceships. *Journal of Abnormal and Social Psychology,* 1955, **51**, 147–148.

Byrne, D., & Wong, T. J. Racial prejudice, interpersonal attraction, and assumed dissimilarity of attitudes. *Journal of Abnormal and Social Psychology,* 1962, **65**, 246–253.

Calvin, A. D., Hoffman, F. K., & Harden, E. L. The effect of intelligence and social atmosphere on group problem solving behavior. *Journal of Social Psychology,* 1957, **45**, 61–74.

Cartwright, D., & Zander, A. *Group dynamics: Research and theory.* Evanston, Ill.: Row, Peterson, 1953.

Cartwright, D., & Zander, A. *Group dynamics: Research and theory.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.

Cervin, V. Individual behavior in social situations: Its relation to anxiety, neuroticism, and group solidarity. *Journal of Experimental Psychology,* 1956, **51**, 161–168.

Chapman, L. J., & Campbell, D. T. An attempt to predict the performance of three-man teams from attitude measurements. *Journal of Social Psychology,* 1957, **46**, 277–286.

Cohen, A. R. Experimental effects of ego-defense preference on interpersonal relations. *Journal of Abnormal and Social Psychology,* 1956, **52**, 19–27.

Cohen, A. R. Upward communication in experimentally created hierarchies. *Human Relations,* 1958, **11**, 41–54.

Cohn, T. S., Yee, W., & Brown, V. Attitude change and interpersonal attraction. *Journal of Social Psychology,* 1961, **55**, 207–211.

Darley, J. G., Gross, N., & Martin, W. E. Studies of group behavior: Factors associated with productivity of groups. *Journal of Applied Psychology,* 1952, **36**, 396–403.

Davitz, J. Social perception and sociometric choice of children. *Journal of Abnormal and Social Psychology,* 1955, **50**, 173–176.

Davol, S. H. An empirical test of structural balance in sociometric triads. *Journal of Abnormal and Social Psychology,* 1959, **59**, 393–398.

de Charms, R. Affiliation motivation and productivity in small groups. *Journal of Abnormal and Social Psychology,* 1957, **55**, 222–226.

Deutsch, M. Some factors affecting membership motivation and achievement motivation in a group. *Human Relations,* 1959, **12**, 81–95.

Deutsch, M. The effects of cooperation and competition upon group process. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 414–448.

Deutsch, M., & Collins, Mary. The effect of public policy in housing projects upon interracial attitudes. In Eleanor Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (3rd ed.) New York: Holt, 1958. Pp. 612–623.

Deutsch, M., & Solomon, L. Reactions to evaluations by others as influenced by self-evaluations. *Sociometry,* 1959, **22**, 93–112.

Dittes, J. E. Attractiveness of group as a function of self-esteem and acceptance by group. *Journal of Abnormal and Social Psychology,* 1959, **59**, 77–82.

Dittes, J. E., & Kelley, H. H. Effects of different conditions of acceptance upon conformity to group norms. *Journal of Abnormal and Social Psychology,* 1956, **53**, 100–107.

Doob, L. W. The behavior of attitudes. *Psychological Review,* 1947, **54**, 135–156.

Downing, J. Cohesiveness, perception, and values. *Human Relations,* 1958, **11**, 157–166.

Eisman, Bernice. Some operational measures of cohesiveness and their interrelations. *Human Relations,* 1959, **12**, 183–189.

Eisman, Bernice, & Levy, J. The influence of certain communicator characteristics on lip reading efficiency. *Journal of Social Psychology,* 1960, **51**, 419–425.

Emerson, R. M. Deviation and rejection: An experimental replication. *American Sociological Review,* 1954, **19**, 688–693.

Ex, J. The nature of contact between co-operating persons and their expectancy concerning the level of their common achievement. *Acta Psychologica,* 1959, **16**, 99–107.

Exline, R. V. Group climate as a factor in the relevance and accuracy of social perception. *Journal of Abnormal and Social Psychology,* 1957, **55**, 382–388.

Exline, R. V., & Ziller, R. C. Status congruency and interpersonal conflict in decision-making groups. *Human Relations,* 1959, **12**, 147–162.

Faunce, D., & Beegle, J. A. Cleavages in a relatively homogeneous group of rural youth. *Sociometry,* 1948, **11**, 207–216.

Festinger, L. Informal social communication. *Psychological Review,* 1950, **57**, 271–282. (a)

Festinger, L. Laboratory experiments: The role of group belongingness. In J. G. Miller (Ed.), *Experiments in social process*. New York: McGraw-Hill, 1950. Pp. 31–46. (b)

Festinger, L. Group attraction and membership. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory*. Evanston, Ill.: Row, Peterson, 1953. Pp. 92–101.

Festinger, L. A theory of social comparison processes. *Human Relations,* 1954, **7**, 117–140.

Festinger, L., & Aronson, E. The arousal and reduction of dissonance in social contexts. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 214–231.

Festinger, L., Gerard, H., Hymovitch, B., Kelley, H., & Raven, B. The influence process in the presence of extreme deviates. *Human Relations,* 1952, **5**, 327–346.

Festinger, L., Schachter, S., & Back, K. *Social pressures in informal groups*. New York: Harper, 1950.

Festinger, L., & Thibaut, J. Interpersonal communication in small groups. *Journal of Abnormal and Social Psychology,* 1951, **46**, 92–99.

Festinger, L., Torrey, Jane, & Willerman, B. Self-evaluation as a function of attraction to the group. *Human Relations,* 1954, **7**, 161–174.

Fey, W. F. Acceptance by others and its relation to acceptance of self and others: A revaluation. *Journal of Abnormal and Social Psychology,* 1955, **50**, 274–276.

Fiedler, F. E. The psychological distance dimension in interpersonal relations. *Journal of Personality,* 1953, **22**, 142–150.

Fiedler, F. E. Assumed similarity measures as predictors of team effectiveness. *Journal of Abnormal and Social Psychology,* 1954, **49**, 381–388.

Fiedler, F. E., Warrington, W. G., & Blaisdell, F. J. Unconscious attitudes as correlates of sociometric choice in a social group. *Journal of Abnormal and Social Psychology,* 1952, **47**, 790–796.

French, J. R. P., Jr. The disruption and cohesion of groups. *Journal of Abnormal and Social Psychology,* 1941, **36**, 361–377.

French, J. R. P., Jr. Group productivity. In H. Guetzkow (Ed.), *Groups, leadership, and men.* Pittsburgh: Carnegie Press, 1951. Pp. 44–54.

French, R. L. Sociometric status and individual adjustment among naval recruits. *Journal of Abnormal and Social Psychology,* 1951, **46**, 64–72.

Gage, N. L., & Exline, R. V. Social perception and effectiveness in discussion groups. *Human Relations,* 1953, **6**, 381–396.

Gage, N. L., & Suci, G. Social perception and teacher-pupil relationships. *Journal of Educational Psychology,* 1951, **42**, 144–152.

Gardner, E., & Thompson, G. *Social relations and morale in small groups*. New York: Appleton-Century-Crofts, 1956.

Gerard, H. B. The anchorage of opinions in face-to-face groups. *Human Relations,* 1954, **7**, 313–325.

Gerard, H. B. Reciprocal effects of the influence upon the influencing agent. Riverside, Calif.: Author, 1961. (Mimeo)

Gilchrist, J. C. The formation of social groups under conditions of success and failure. *Journal of Abnormal and Social Psychology,* 1952, **47**, 174–187.

Goodacre, D. M. The use of a sociometric test as a predictor of combat unit effectiveness. *Sociometry,* 1951, **14**, 148–152.

Goodacre, D. M. Group characteristics of good and poor performing combat units. *Sociometry,* 1953, **16**, 168–178.

Goodnow, R. E., & Tagiuri, R. Religious ethnocentrism and its recognition among adolescent boys. *Journal of Abnormal and Social Psychology,* 1952, **47**, 316–320.

Gottheil, E. Changes in social perceptions contingent upon competing or cooperating. *Sociometry,* 1955, **18**, 132–137.

Greer, F. L., Galanter, E. H., & Nordlier, P. G. Interpersonal knowledge and individual and group effectiveness. *Journal of Abnormal and Social Psychology,* 1954, **49**, 411–414.

Gross, E. Primary functions of the small group. *American Journal of Sociology,* 1954, **60**, 24–29.

Gross, E. Symbiosis and consensus as integrative factors in small groups. *American Sociological Review,* 1956, **21**, 174–179.

Grossack, M. M. Some effects of cooperation and competition upon small group behavior. *Journal of Abnormal and Social Psychology,* 1954, **49**, 341–348.

Grosser, D., Polansky, N., & Lippitt, R. A laboratory study of behavior contagion. *Human Relations,* 1951, **4**, 115–142.

Guetzkow, H., & Gyr, J. An analysis of conflict in decision-making groups. *Human Relations,* 1954, **7**, 367–382.

Gullahorn, J. Distance and friendship as factors in the gross interaction matrix. *Sociometry,* 1952, **15**, 123–134.

Gundlach, R. H. Effects of on-the-job experiences with Negroes upon racial attitudes of white workers in union shops. *Psychological Reports,* 1956, **2**, 67–77.

Gurnee, H. Group learning. *Psychological Monographs,* 1962, **76**(13, Whole No. 532).

Hamblin, R. L. Group integration during a crisis. *Human Relations,* 1958, **11**, 67–76.

Harvey, O. J. Personality factors in resolution of conceptual incongruities. *Sociometry,* 1962, **25**, 336–352.

Harvey, O. J., & Consalvi, C. Status and conformity to pressures in informal groups. *Journal of Abnormal and Social Psychology,* 1960, **60**, 182–187.

Harvey, O. J., Kelley, H. H., & Shapiro, M. M. Reactions to unfavorable evaluations of the self made by other persons. *Journal of Personality,* 1957, **25**, 393–411.

Haythorn, W. The influence of individual members on the characteristics of small groups. *Journal of Abnormal and Social Psychology,* 1953, **48**, 276–284.

Heber, R. F., & Heber, Mary E. The effect of group failure and success on social status. *Journal of Educational Psychology,* 1957, **48**, 129–134.

Heider, F. Attitudes and cognitive organization. *Journal of Psychology,* 1946, **21**, 107–112.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Hoffman, L. R. Similarity of personality: A basis for interpersonal attraction? *Sociometry,* 1958, **21**, 300–308.

Hoffman, L. R. Homogeneity of member personality and its effect on group problem-solving. *Journal of Abnormal and Social Psychology,* 1959, **58**, 27–32.

Hoffman, L. R., & Maier, N. R. F. Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *Journal of Abnormal and Social Psychology,* 1961, **62**, 401–407.

Homans, G. C. *Social behavior: Its elementary forms.* New York: Harcourt, Brace, 1961.

Horowitz, M. W., Lyons, J., & Perlmutter, H. V. Induction of forces in discussion groups. *Human Relations,* 1951, **4**, 57–76.

Horsfall, A. B., & Arensberg, C. M. Teamwork and productivity in a shoe factory. *Human Organization,* 1949, **8**(1), 13–25.

Horwitz, M. The veridicality of liking and disliking. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 191–209.

Howard, R. C., & Berkowitz, L. Reactions to the evaluators of one's performance. *Journal of Personality,* 1958, **26**, 494–507.

Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

Hurwitz, J. I., Zander, A. F., & Hymovitch, B. Some effects of power on the relations among group members. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 800–809.

Husband, R. W. Cooperative versus solitary problem solution. *Journal of Social Psychology,* 1940, **11**, 405–409.

Izard, C. E. Personality similarity and friendship. *Journal of Abnormal and Social Psychology,* 1960, **61**, 47–51. (a)

Izard, C. E. Personality similarity, positive affect, and interpersonal attraction. *Journal of Abnormal and Social Psychology,* 1960, **61**, 484–485. (b)

Jackson, J. M. Reference group processes in a formal organization. *Sociometry,* 1959, **22**, 307–327.

James, Anna, & Lott, A. J. Reward frequency and the formation of positive attitudes toward group members. *Journal of Social Psychology,* 1964, **62**, 111–115.

Jennings, Helen H. *Leadership and isolation.* (2nd ed.) New York: Longmans, Green, 1950.

Jones, E. E., & Daugherty, B. N. Political orientation and the perceptual effects of an anticipated interaction. *Journal of Abnormal and Social Psychology,* 1959, **59**, 340–349.

Jordan, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Human Relations,* 1953, **6**, 273–287.

Kanfer, F. H., & Karas, Shirley C. Prior experimenter-subject interaction and verbal conditioning. *Psychological Reports,* 1959, **5**, 345–353.

Kates, S. L. First-impression formation and authoritarianism. *Human Relations,* 1959, **12**, 277–286.

Katz, I., & Cohen, M. The effects of training Negroes upon cooperative problem solving in biracial teams. *Journal of Abnormal and Social Psychology,* 1962, **64**, 319–325.

Kelley, H. H. The warm-cold variable in first impressions of persons. *Journal of Personality,* 1950, **18**, 431–439.

Kelley, H. H. Communication in experimentally created hierarchies. *Human Relations,* 1951, **4**, 39–56.

Kelley, H. H., & Shapiro, M. M. An experiment on conformity to group norms where conformity is detrimental to group achievement. *American Sociological Review,* 1954, **19**, 667–677.

Kelley, H. H., & Thibaut, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass., Addison-Wesley, 1954. Pp. 735–785.

Kelley, H. H., & Volkart, E. H. The resistance to change of group-anchored attitudes. *American Sociological Review,* 1952, **17**, 453–465.

Kelley, H. H., & Woodruff, Christine L. Members' reactions to apparent group approval of a counter-norm communication. *Journal of Abnormal and Social Psychology,* 1956, **52**, 67–74.

Kidd, J. S., & Campbell, D. T. Conformity to groups as a function of group success. *Journal of Abnormal and Social Psychology,* 1955, **51**, 390–393.

Kipnis, Dorothy M. Interaction between members of bomber crews as a determinant of sociometric choice. *Human Relations,* 1957, **10**, 263–270.

Kipnis, Dorothy M. Changes in self concepts in relation to perceptions of others. *Journal of Personality,* 1961, **29**, 449–465.

Kleiner, R. J. The effects of threat reduction upon interpersonal attraction. *Journal of Personality,* 1960, **28**, 145–155.

Kogan, N., & Tagiuri, R. Interpersonal preference and cognitive organization. *Journal of Abnormal and Social Psychology,* 1958, **56**, 113–116.

Lanzetta, J. T. Group behavior under stress. *Human Relations,* 1955, **8**, 29–52.

Lanzetta, J. T., Haefner, D., Langham, P., & Axelrod, H. Some effects of situational threat on group behavior. *Journal of Abnormal and Social Psychology,* 1954, **3**, 445–453.

Lazarsfeld, P. F., & Merton, R. K. Friendship as social process: A substantive and methodological analysis. In M. Berger, T. Abel, & C. H. Page (Eds.), *Freedom and control in modern society.* New York: Van Nostrand, 1954. Pp. 18–66.

Lemann, T. B., & Solomon, R. L. Group characteristics as revealed in sociometric patterns and personality ratings. *Sociometry,* 1952, **15**, 7–90.

Lerner, M. J., & Becker, S. Interpersonal choice as a function of ascribed similarity and definition of the situation. *Human Relations,* 1962, **15**, 27–34.

Libo, L. *Measuring group cohesiveness.* Ann Arbor: Univer. Michigan Press, 1953. (Institute for Social Research, Research Center for Group Dynamics)

Libo, L. The projective expression of patient-therapist attraction. *Journal of Clinical Psychology,* 1957, **13**, 33–36.

Lindzey, G., & Borgatta, E. F. Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954. Pp. 405–448.

Lindzey, G., & Urdan, J. A. Personality and social choice. *Sociometry,* 1954, **17**, 47–63.

Lippitt, R., Polansky, N., & Rosen, S. The dynamics of power. *Human Relations,* 1952, **5**, 37–64.

Lott, A. J., & Lott, Bernice E. Group cohesiveness, communication level and conformity. *Journal of Abnormal and Social Psychology,* 1961, **62**, 408–412.

Lott, Bernice E. Group Cohesiveness: A learning phenomenon. *Journal of Social Psychology,* 1961, **55**, 275–286.

Lott, Bernice E., & Lott, A. J. The formation of positive attitudes toward group members. *Journal of Abnormal and Social Psychology,* 1960, **61**, 297–300.

Lorge, I., Fox, D., Davitz, J., & Brenner, Marilyn. A survey of studies contrasting the quality of group performance and individual performance, 1920–1957. *Psychological Bulletin,* 1958, **55**, 337–372.

Lundy, R. M. The relationship of changes in assimilative projection to accepting and rejecting interpersonal groups and to the order of the groups. *Journal of Social Psychology,* 1959, **50**, 327–333.

Lundy, R. M., Katkovsky, W., Cromwell, R. L., & Shoemaker, D. J. Self acceptability and descriptions of sociometric choices. *Journal of Abnormal and Social Psychology,* 1955, **51**, 260–262.

Maehr, M. L., Mensing, J., & Nafzger, S. Concept of self and the reaction of others. *Sociometry,* 1962, **25**, 353–357.

Maissonneuve, J. A contribution to the sociometry of mutual choices. *Sociometry,* 1954, **17**, 33–46.

Maissonneuve, J., Palmade, G., & Fourment, C. Selective choices and propinquity. *Sociometry,* 1952, **15**, 135–140.

Malof, M., & Lott, A. J. Ethnocentrism and the acceptance of Negro support in a group pressure situation. *Journal of Abnormal and Social Psychology,* 1962, **65**, 254–258.

Manis, M. Social interaction and the self concept. *Journal of Abnormal and Social Psychology,* 1955, **51**, 362–370.

Mann, J. H. The effect of inter-racial contact on sociometric choices and perceptions. *Journal of Social Psychology,* 1959, **50**, 143–152.

Mann, J. H., & Mann, C. H. The importance of a group task in producing group member personality and behavior changes. *Human Relations,* 1959, **12**, 75–80.

Mann, R. D. A review of the relationships between personality and performance in small groups. *Psychological Bulletin,* 1959, **56**, 241–270.

Marks, J. B. Interests and group formation. *Human Relations,* 1959, **12**, 385–390.

Marquis, D. G., Guetzkow, H., & Heyns, R. W. A social psychological study of the decision-making conference. In H. Guetzkow (Ed.), *Groups, leadership, and men.* Pittsburgh: Carnegie Press, 1951. Pp. 55–67.

Masling, J., Greer, F. L., & Gilmore, R. Status, authoritarianism and sociometric choice. *Journal of Social Psychology,* 1955, **41**, 297–310.

McDavid, J. W. Imitative behavior in preschool children. *Psychological Monographs,* 1959, **73**(16, Whole No. 486).

McGrath, J. E. The influence of positive interpersonal relations on adjustment and effectiveness in rifle teams. *Journal of Abnormal and Social Psychology,* 1962, **65**, 365–375.

McIntyre, C. J. Acceptance by others and its relation to acceptance of self and others. *Journal of Abnormal and Social Psychology,* 1952, **47**, 624–625.

McKeachie, W. J. Individual conformity to attitudes of classroom groups. *Journal of Abnormal and Social Psychology,* 1954, **49**, 282–289.

Mellinger, G. D. Interpersonal trust as a factor in communication. *Journal of Abnormal and Social Psychology,* 1956, **52**, 304–309.

Miller, D. R. The study of social relationships: Situation identity and social interaction. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 5. New York: McGraw-Hill, 1963. Pp. 639–737.

Miyamoto, S. F., & Dornbusch, S. M. A test of interactionist hypotheses of self-conception. *American Journal of Sociology,* 1956, **61**, 399–403.

Moos, R. H., & Speisman, J. C. Group compatibility and productivity. *Journal of Abnormal and Social Psychology,* 1962, **65**, 190–196.

Myers, A. Team competition, success, and the adjustment of group members. *Journal of Abnormal and Social Psychology,* 1962, **65**, 325–332.

Newcomb, T. M. An approach to the study of communicative acts. *Psychological Review,* 1953, **60**, 393–404.

Newcomb, T. M. The prediction of interpersonal attraction. *American Psychologist,* 1956, **11**, 575–586.

Newcomb, T. M. The cognition of persons as cognizers. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 179–190.

Newcomb, T. M. Varieties of interpersonal attraction. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 104–119.

Newcomb, T. M. *The acquaintance process.* New York: Holt, Rinehart & Winston, 1961.

Norfleet, B. Interpersonal relations and group productivity. *Journal of Social Issues,* 1948, **4**, 66–69.

Palmore, E. B. The introduction of Negroes into white departments. *Human Organization,* 1955, 14(1) 27–28.

Pepinsky, Pauline N., Hemphill, J. K., & Shevitz, R. N. Attempts to lead, group productivity, and morale under conditions of acceptance and rejection. *Journal of Abnormal and Social Psychology,* 1958, **57**, 47–54.

Pepitone, A. Attributions of causality, social attitudes, and cognitive matching processes. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 258–276.

Pepitone, A., & Kleiner, R. The effects of threat and frustration on group cohesiveness. *Journal o Abnormal and Social Psychology,* 1957, **54**, 192–199.

Pepitone, A., & Reichling, G. Group cohesiveness and the expression of hostility. *Human Relations* 1955, **8**, 327–337.

Pepitone, A., & Sherberg, Janet. Intentionality responsibility, and interpersonal attraction. *Journal of Personality,* 1957, **25**, 757–766.

Pepitone, A., & Wilpizeski, C. Some consequences of experimental rejection. *Journal of Abnormal and Social Psychology,* 1960, **60**, 359–364.

Perkins, H. V., Jr. The effects of climate and curriculum on group learning. *Journal of Educational Research,* 1950, **44**, 269–286.

Perlmutter, H. V. Impressions of influential members of discussion groups. *Journal of Psychology,* 1954, **38**, 223–234.

Phillips, B. N., & D'Amico, L. A. Effects of cooperation and competition on the cohesiveness of small face-to-face groups. *Journal of Educational Psychology,* 1956, **47**, 65–70.

Philp, Alice J. Strangers and friends as competitors and cooperators. *Journal of Genetic Psychology,* 1940, **57**, 249–258.

Pilisuk, M. Cognitive balance and self-relevant attitudes. *Journal of Abnormal and Social Psychology,* 1962, **65**, 95–103.

Precker, J. A. Similarity of valuings as a factor in selection of peers and near authority figures. *Journal of Abnormal and Social Psychology,* 1952, **47**, 406–414.

Precker, J. A. The automorphic process in the attribution of values. *Journal of Personality,* 1953, **21**, 356–363.

Preston, M. G., Peltz, W. L., Mudd, E. H., & Froscher, H. B. Impressions of personality as a function of marital conflict. *Journal of Abnormal and Social Psychology,* 1952, **47**, 326–336.

Ramuz-Nienhuis, Wilhelmina, & van Bergen, Annie. Relations between some components of attraction-to-group: A replication. *Human Relations,* 1960, **13**, 271–277.

Rasmussen, G., & Zander, A. Group membership and self-evaluation. *Human Relations,* 1954, **7**, 239–251.

Raven, B. H., & French, J. R. P., Jr. Group support, legitimate power, and social influence. *Journal of Personality,* 1958, **26**, 400–409.

Reece, M. M., & Whitman, R. N. Expressive movements, warmth, and verbal reinforcement. *Journal of Abnormal and Social Psychology,* 1962, **64**, 234–236.

Reese, H. W. Relationships between self-acceptance and sociometric choices. *Journal of Abnormal and Social Psychology,* 1961, **62**, 472–474.

Rehage, K. J. A comparison of pupil-teacher planning and teacher-directed procedures in eighth grade social studies classes. *Journal of Educational Research,* 1951, **45**, 111–115.

Reilly, Mary St. Anne, Commins, W. D., & Stefic, E. C. The complementarity of personality needs in friendship choice. *Journal of Abnormal and Social Psychology*, 1960, **61**, 292–294.

Riecken, H. W., & Homans, G. C. Psychological aspects of social structure. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 786–832.

Rokeach, M. *The open and closed mind*. New York: Basic Books, 1960.

Roseborough, Mary. Experimental studies of small groups. *Psychological Bulletin*, 1953, **50**, 275–303.

Rosenbaum, M. E. Social perception and the motivational structure of interpersonal relations. *Journal of Abnormal and Social Psychology*, 1959, **59**, 130–133.

Runkel, P. J. Replicated tests of the attraction-communication hypothesis in a setting of technical information flow. *American Sociological Review*, 1962, **27**, 402–408.

Sachs, Elinor L. Intelligence scores as a function of experimentally established relationships between child and examiner. *Journal of Abnormal and Social Psychology*, 1952, **47**, 354–358.

Sapolsky, A. Effect of interpersonal relationships upon verbal conditioning. *Journal of Abnormal and Social Psychology*, 1960, **60**, 241–246.

Schachter, S. Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology*, 1951, **46**, 190–207.

Schachter, S., Ellertson, N., McBride, Dorothy, & Gregory, Doris. An experimental study of cohesiveness and productivity. *Human Relations*, 1951, **4**, 229–238.

Schachter, S., Nuttin, J., De Monchaux, C., Maucorps, P., Osmer, D., Duijker, H., Rommetveit, R., & Israel, J. Cross-cultural studies on threat and rejection. *Human Relations*, 1954, **7**, 403–439.

Schopler, J., & Bateson, N. A dependence interpretation of the effects of a severe initiation. *Journal of Personality*, 1962, **30**, 633–649.

Schutz, W. C. What makes groups productive? *Human Relations*, 1955, **8**, 429–465.

Schutz, W. C. *FIRO: A three-dimensional theory of interpersonal behavior*. New York: Rinehart, 1958.

Schutz, W. C. On group composition. *Journal of Abnormal and Social Psychology*, 1961, **62**, 275–281.

Seashore, S. E. *Group cohesiveness in the industrial work group*. Ann Arbor: Univer. Michigan Press, 1954. (Institute of Social Relations, Survey Research Center)

Shaw, M. E. Group dynamics. *Annual Review of Psychology*, 1961, **12**, 129–156.

Shaw, M. E., & Gilchrist, J. C. Repetitive task failure and sociometric choice. *Journal of Abnormal and Social Psychology*, 1955, **50**, 29–32.

Shaw, M. E., & Shaw, Lilly May. Some effects of sociometric grouping upon learning in a second grade classroom. *Journal of Social Psychology*, 1962, **57**, 453–458.

Shelley, H. P. Level of aspiration phenomena in small groups. *Journal of Social Psychology*, 1954, **40**, 149–164.

Shelley, H. P. Focused leadership and cohesiveness in small groups. *Sociometry*, 1960, **23**, 209–216.

Sherif, M., & Sherif, Carolyn. *Groups in harmony and tension*. New York: Harper, 1953.

Sherif, M., White, B. J., & Harvey, O. J. Status in experimentally produced groups. *American Journal of Sociology*, 1955, **60**, 370–379.

Simkins, L. Effects of examiner attitudes and type of reinforcement on the conditioning of hostile verbs. *Journal of Personality*, 1961, **29**, 380–395.

Smith, A. J. Similarity of values and its relation to acceptance and the projection of similarity. *Journal of Psychology*, 1957, **43**, 251–260.

Smith, A. J. Perceived similarity and the projection of similarity: The influence of valence. *Journal of Abnormal and Social Psychology*, 1958, **57**, 376–379.

Smith, A. J., Madden, H. E., & Sobol, R. Productivity and recall in cooperative and competitive discussion groups. *Journal of Psychology*, 1957, **43**, 193–204.

Smith, E. E. The effects of clear and unclear role expectations on group productivity and defensiveness. *Journal of Abnormal and Social Psychology*, 1957, **55**, 213–217.

Snoek, J. D. Some effects of rejection upon attraction to a group. *Journal of Abnormal and Social Psychology*, 1962, **64**, 175–182.

Solomon, L. The influence of some types of power relationships and game strategies upon the development of interpersonal trust. *Journal of Abnormal and Social Psychology*, 1960, **61**, 223–230.

Spector, A. J. Expectations, fulfillment, and morale. *Journal of Abnormal and Social Psychology*, 1956, **52**, 51–56.

Speroff, E., & Kerr, W. Steel mill "hot strip" accidents and interpersonal desirability values. *Journal of Clinical Psychology*, 1952, **8**, 89–91.

Steiner, I. D., & Dodge, Joan. Interpersonal perception and role structure as determinants of group and individual efficiency. *Human Relations*, 1956, **9**, 467–480.

Stendler, Celia, Damrin, Dora, & Haines, Aleyne. Studies in cooperation and competition: I. The effects of working for group and individual rewards on the social climate of children's groups. *Journal of Genetic Psychology*, 1951, **79**, 173–197.

Stogdill, R. *Individual behavior and group achievement: A theory, the experimental evidence*. New York: Oxford Univer. Press, 1959.

Stotland, E. Determinants of attraction to groups. *Journal of Social Psychology*, 1959, **49**, 71–80.

Stotland, E., & Cottrell, N. B. Similarity of performance as influenced by interaction, self-esteem, and birth order. *Journal of Abnormal and Social Psychology*, 1962, **64**, 183–191.

Stotland, E., Cottrell, N. B., & Laing, G. Group interaction and perceived similarity of members. *Journal of Abnormal and Social Psychology*, 1960, **61**, 335–340.

Stotland, E., & Hillmer, M. L., Jr. Identification, authoritarian defensiveness, and self-esteem. *Journal of Abnormal and Social Psychology*, 1962, **64**, 334–342.

Stotland, E., Zander, A., & Natsoulas, T. Generalization of interpersonal similarity. *Journal of Abnormal and Social Psychology*, 1961, **62**, 250–256.

Suchman, J. R. Social sensitivity in the small task-oriented group. *Journal of Abnormal and Social Psychology*, 1956, **52**, 75–83.

Tagiuri, R. Relational analysis: An extension of sociometric methods with emphasis upon social perception. *Sociometry*, 1952, **15**, 91–104.

Tagiuri, R. Social preference and its perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 316–336.

Tagiuri, R., Blake, R. R., & Bruner, J. S. Some determinants of the perception of positive and negative feelings in others. *Journal of Abnormal and Social Psychology*, 1953, **48**, 585–592.

Tagiuri, R., Bruner, J. S., & Blake, R. R. On the relation between feelings and perception of feelings among members of small groups. In Eleanor Maccoby, T. Newcomb, & E. Hartley (Eds.), *Readings in social psychology*. (3rd ed.) New York: Holt, 1958. Pp. 110–116.

Tagiuri, R., & Kogan, N. Personal preference and the attribution of influence in small groups. *Journal of Personality*, 1960, **28**, 257–265.

Tagiuri, R., Kogan, N., & Long, L. M. Differentiation of sociometric choice and status relations in a group. *Psychological Reports*, 1958, **4**, 523–526.

Theodorson, G. A. The relationship between leadership and popularity roles in small groups. *American Sociological Review*, 1957, **22**, 58–67.

Thibaut, J. W. An experimental study of the cohesiveness of underprivileged groups. *Human Relations*, 1950, **3**, 251–278.

Thibaut, J. W., & Kelley, H. H. *The social psychology of groups*. New York: Wiley, 1959.

Thibaut, J. W., & Riecken, H. W. Authoritarianism, status, and the communication of aggression. *Human Relations*, 1955, **8**, 95–120.

Thibaut, J. W., & Strickland, L. H. Psychological set and social conformity. *Journal of Personality*, 1956, **25**, 115–129.

Thompson, W. R., & Nishimura, Rhoda. Some determinants of friendship. *Journal of Personality*, 1952, **20**, 305–314.

Thrasher, J. D. Interpersonal relations and gradations of stimulus structure as factors in judgmental variation: An experimental approach. *Sociometry*, 1954, **17**, 228–241.

Triandis, H. C. A note on Rokeach's theory of prejudice. *Journal of Abnormal and Social Psychology*, 1961, **62**, 184–186.

Van Bergen, Annie, & Koekebakker, J. "Group cohesiveness" in laboratory experiments. *Acta Psychologica*, 1959, **16**, 81–98.

Van Zelst, R. H. Worker popularity and job satisfaction. *Personnel Psychology*, 1951, **4**, 405–412.

Van Zelst, R. H. Sociometrically selected work teams increase production. *Personnel Psychology*, 1952, **5**, 175–185. (a)

Van Zelst, R. H. Validation of a sociometric regrouping procedure. *Journal of Abnormal and Social Psychology*, 1952, **47**, 299–301. (b)

Videbeck, R. Self-conception and the reactions of others. *Sociometry*, 1960, **23**, 351–359.

Walker, E. L., & Heyns, R. W. *An anatomy for conformity*. Englewood Cliffs, N. J.: Prentice-Hall, 1962.

Weiss, R. L., Krasner, L., & Ullmann, L. P. Responsivity to verbal conditioning as a function of emotional atmosphere and pattern of reinforcement. *Psychological Reports*, 1960, **6**, 415–426.

White, R., & Lippitt, R. Leader behavior and member reaction in three "social climates." In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 527–553.

Willerman, B., & Swanson, L. An ecological determinant of differential amounts of sociometric choices within college sororities. *Sociometry*, 1952, **15**, 326–329.

Wilson, W., & Miller, N. Shifts in evaluations of participants following intergroup competition. *Journal of Abnormal and Social Psychology*, 1961, **63**, 428–431.

Winch, R. F. The theory of complementary needs in mate selection: A test of one kind of complementariness. *American Sociological Review*, 1955, **20**, 52–56.

Winch, R. F., Ktanes, T., & Ktanes, Virginia. Empirical elaboration of the theory of complementary needs in mate selection. *Journal of Abnormal and Social Psychology*, 1955, **51**, 508–513.

Wright, M. E. The influence of frustration upon the social relations of young children. *Character and Personality*, 1943, **12**, 111–122.

Yuker, H. E. Group atmosphere and memory. *Journal of Abnormal and Social Psychology*, 1955, **51**, 17–23.

Zander, A., & Cohen, A. R. Attributed social power and group acceptance: A classroom experimental demonstration. *Journal of Abnormal and Social Psychology*, 1955, **51**, 490–492.

Zander, A., & Havelin, A. Social comparison and interpersonal attraction. *Human Relations*, 1960, **13**, 21–32.

Zander, A., Natsoulas, T., & Thomas, E. J. Personal goals and group's goals for the members. *Human Relations*, 1960, **13**, 333–344.

Zander, A., Stotland, E., & Wolfe, D. Unity of group, identification with group, and self esteem of members. *Journal of Personality*, 1960, **28**, 463–478.

Ziller, R. C., & Behringer, R. D. Assimilation of the knowledgeable newcomer under conditions of group success and failure. *Journal of Abnormal and Social Psychology*, 1960, **60**, 288–291.

Zimmer, H. Motivational factors in dyadic interaction. *Journal of Personality*, 1956, **24**, 251–261.

(Received April 28, 1965)

Psychological Bulletin
1965, Vol. 64, No. 4, 310–312

REVERSAL LEARNING AND RESISTANCE TO EXTINCTION:

A SUPPLEMENTARY REPORT

SALLY E. SPERLING
University of California, Riverside

Several very recent and highly relevant articles provide substantiation for some hypotheses advanced in a previous review in an analysis of the overtraining reversal data. These same data permit the rejection of an hypothesis advanced by another reviewer to account for the results. It can now be shown fairly conclusively that amount of reward determines the direction of the ORE following position response training and that the variables affecting overtraining reversal data with exteroceptive discriminative stimuli remain unidentified.

In the time between acceptance and publication of a previous literature review (Sperling, 1965), a paper by Paul (1965) on the overtraining reversal effect (ORE) was published. This article cited several relevant experiments that had not been discussed in the Sperling article and advanced a suggestion for accounting for the differences in results obtained in comparing the reversal performance of rats after training to criterion on the initial problem with that of rats given overtraining on the initial problem before reversal. During the same period three extremely relevant papers were made available to the writer (D'Amato & Schiff, 1965; Theios & Blosser, 1965a, 1965b), all of which have now appeared in print.

This accumulation of articles permitted the substantiation of some hypotheses advanced in the analysis of the ORE data in Sperling (1965) and the rejection of the hypothesis advanced by Paul (1965) to account for these data. These seemed sufficient reasons to relate these new data in this supplementary report to those data previously discussed.

Additions to the Experiments Using Exteroceptive Discriminative Stimulus Training. D'Amato and Schiff (1965) used light-intensity differences as discriminative stimuli in a series of eight experiments in an automated Y maze. (This is the same apparatus in which D'Amato & Jagoda, 1961, observed the ORE.) They manipulated a substantial number of variables—degree of stimulus change from acquisition to reversal, intertrial interval (ITI), amount of reward, strain of rats, presence or absence of discriminanda during ITI, and variations designed to affect the rate of extinction of the initial habit—and found no evidence of an ORE in the single studies or in the data of all of them combined. (It should perhaps be noted that the data on reversal learning using exteroceptive discriminative stimuli show either an ORE or no evidence of an ORE. To the writer's knowledge, a reverse ORE has not been observed for this kind of training.)

The present writer shares the confusion and frustration expressed by the experimenters; as in the case of Erlebacher (1963), there is no discernible reason for the lack of an ORE in these data. The results of the D'Amato and Schiff (1965) and Erlebacher (1963) experiments must lead to the conclusion that the ORE is not a general phenomenon even for exteroceptive discriminative stimulus training in rats. The set of conditions under which it can be expected to occur must, therefore, be at least somewhat limited and specific (especially since, as mentioned by the investigators, the conditions used in these experiments were present in one combination or another in other experiments showing an ORE). There seems to be no obvious hint in these data about what the specific conditions may be. Nevertheless, it is the case that an ORE is observed about half of the time following exteroceptive discriminative stimulus training, and it seems unlikely that its occurrence or nonoccurrence is due to chance.

Additions to the Experiments Using Spatial or Position Response Training. Both Pubols (1956, Experiment II) and Capaldi (1963) observed an ORE following position response training. Initial acquisition was quite rapid in these two studies compared to other position response experiments. Subjects began their criterion run in an average of five trials in the first study and were responding with over 90% correct choices after 16 trials in the second study. It is difficult to see, however, how this would differentially affect the reversal data.

Sperling (1965) suggested that amount of reward is the crucial variable affecting the kind of overtraining reversal data observed following position response training. If, as was hypothesized, the relationship between the resistance to extinction of the previously reinforced response and amount or magnitude of reward is negative or nonmonotonic, a reverse ORE, no evidence of an ORE, and finally an ORE would be predicted as amount of reward increased. This relationship seems to explain the above data.

Unfortunately, Pubols (1956) specified neither the size nor the weight of his "one reinforcement pellet," but the comparably rapid acquisition suggests that it was, as in Capaldi (1963), relatively large. In at least the latter experiment, the amount of reward was large (judging from the 15-second access time) which undoubtedly accounted for the observation of the ORE.

Two experiments by Theios and Blosser (1965a, 1965b) fairly conclusively substantiate the hypothesis about the effect of amount of reward on reversal following position response training. They used both a nondifferentiated T maze and one less differentiated than that in Kendler and Kimm (1964). In three experiments in the two reports, an ORE was observed for large reward groups, and no evidence of an ORE or a reverse ORE was observed in different small reward groups. The difference between the small and large amounts of reward in these experiments was great enough, therefore, to show both of the end effects of variations in this variable suggested in Sperling (1965).

This same variable also provides a tentative explanation of the observation of the ORE in the 12-hour deprived group in Bruner, Mandler, O'Dowd, and Wallach (1958) and the lack of an ORE in the 36-hour group. If it can be assumed that the eating rate was no faster for the 36-hour than for the 12-hour deprived subjects during the fixed-access-to-food interval used as the reinforcer, then it is perhaps conceivable that the amount consumed constituted a larger reward on some dimension for the less deprived than for the more deprived subjects, since it perhaps brought them closer to satiation. Although these speculations are not testable in the data, they provide an account of them that is consistent with the effects of amount of reward.

What constitutes a large or a small reward in any experimental situation is, of course, going to depend upon the deprivation level, experimental apparatus and response required, training procedures, and perhaps other factors as well, but the effect of amount of reward, already observable at criterion performance on the initial problem (cf. Sperling, 1965), seems without question to be further accentuated by increasing amounts of this kind of training.

Comments on the Effects of Correction and Noncorrection Training Procedures. Paul (1965) suggested that the lack of agreement in results among the ORE data may be related to differences in the experimental sequence following an incorrect response. Experiments both showing and failing to show an ORE used noncorrection procedures, but there was a difference in how the procedures were applied. In most of the former studies an incorrect response led to removal from the apparatus without exposure to the unbaited goal box, while in most of the latter studies an incorrect response led to detention in the unbaited negative goal box. He speculated that this difference in the experimental sequence following an incorrect response may account for the differences in results in the overtraining reversal data even though a few experiments reported an ORE when an incorrect response led to an unbaited goal box. It can be demonstrated that this hypothesis does not aid in explaining the data following either position response training or training using exteroceptive discriminative stimuli alone; consequently, it cannot handle the combined results either.

In position response training, Theios and Blosser (1965a) used noncorrection detention for 5 seconds on error trials and observed an ORE for large reward and no evidence of an ORE for small reward groups in the same experiment. In Theios and Blosser (1965b), both a standard correction procedure and noncorrection detention were used; an ORE was observed for large reward and no ORE or a reverse ORE for small reward groups. These two experiments, then, fairly conclusively show that the effect of overtraining on reversal learning in this situation is predictable from a knowledge of the amount of reward irrespective of the use of either correction or noncorrection procedures.

The results of studies using exteroceptive discriminative stimulus situations also do not seem to depend upon this procedural variation. Neither D'Amato and Schiff (1965) nor Erlebacher (1963) observed the ORE, and both used noncorrection detention in an unbaited goal box on incorrect trials. Among others, Birch, Ison, and Sperling (1960), Capaldi and Stevenson (1957), and Mackintosh (1962, 1963a, 1963b) also used this procedure, and all observed the ORE. (The experiments by Mackintosh—1962, 1963a, 1963b—are counted in the noncorrection detention studies since subjects were detained for 10 seconds in the presence of the negative discriminandum on incorrect choice trials, even though this detention was on a platform in front of the stimulus rather than on a platform behind it.)

SUMMARY

All of these additional data suggest even more strongly the necessity and usefulness of the distinction between training using spatial or position responses and training using exteroceptive discriminative stimuli in analyzing the data on the effects of overtraining on reversal learning. In the former instance, there seems little doubt that the effective variable has now been isolated; in the latter instance, there seems little doubt that it has (or they have) not, and that none of the more obvious avenues of analysis seems to be leading to fruitful hypotheses.

REFERENCES

Birch, D., Ison, J. R., & Sperling, Sally E. Reversal learning under single stimulus presentation. *Journal of Experimental Psychology,* 1960, **60,** 36–40.

Bruner, J. S., Mandler, Jean M., O'Dowd, D., & Wallach, M. A. The role of overlearning and drive level in reversal learning. *Journal of Comparative and Physiological Psychology,* 1958, **51,** 607–613.

Capaldi, E. J. Overlearning reversal effect in a spatial discrimination test. *Perceptual and Motor Skills,* 1963, **16,** 335–336.

Capaldi, E. J., & Stevenson, H. W. Response reversal following different amounts of training. *Journal of Comparative and Physiological Psychology,* 1957, **50,** 195–198.

D'Amato, M. R., & Jagoda, H. Analysis of the role of overlearning in discrimination reversal. *Journal of Experimental Psychology,* 1961, **61,** 45–50.

D'Amato, M. R., & Schiff, D. Overlearning and brightness discrimination reversal. *Journal of Experimental Psychology,* 1965, **69,** 375–381.

Erlebacher, A. Reversal learning in rats as a function of percentage of reinforcement and degree of learning. *Journal of Experimental Psychology,* 1963, **66,** 84–90.

Kendler, H. H., & Kimm, J. Reinforcement and cue factors in reversal learning. *Psychonomic Science,* 1964, **1,** 309–310.

Mackintosh, N. J. The effects of overtraining on a reversal and a nonreversal shift. *Journal of Comparative and Physiological Psychology,* 1962, **55,** 555–559.

Mackintosh, N. J. The effect of irrelevant cues on reversal learning in the rat. *British Journal of Psychology,* 1963, **54,** 127–134. (a)

Mackintosh, N. J. Extinction of a discrimination habit as a function of overtraining. *Journal of Comparative and Physiological Psychology,* 1963, **56,** 842–847. (b)

Paul, C. Effects of overlearning upon single habit reversal in rats. *Psychological Bulletin,* 1965, **63,** 65–72.

Pubols, B. H., Jr. The facilitation of visual and spatial discrimination reversal by overtraining. *Journal of Comparative and Physiological Psychology,* 1956, **49,** 243–248.

Sperling, Sally E. Reversal learning and resistance to extinction: A review of the rat literature. *Psychological Bulletin,* 1965, **63,** 281–297.

Theios, J., & Blosser, D. An incentive model for the overlearning reversal effect. *Psychonomic Science,* 1965, **2,** 37–38. (a)

Theios, J., & Blosser, D. The overlearning reversal effect and magnitude of reward. *Journal of Comparative and Physiological Psychology,* 1965, **59,** 252–257. (b)

(Received February 16, 1965)

Vol. 64, No. 5 November 1965

Psychological Bulletin

SIGNIFICANCE OF ALL-OR-NONE LEARNING[1]

FRANK RESTLE

Indiana University

The essential characteristic of all-or-none data is that errors are recurrent events. Theoretical models having this property are derived from cognitive theory, Hullian theory, and stimulus-sampling theory. In addition, continuous-learning models, including the linear model and Luce's beta model, are shown to have all-or-none special cases. 3 experimental methods—the substitution method, the miniature experiment, and detailed distributional analysis of simple learning—have yielded information on conditions that give all-or-none data. Multiple-stage models, built of more than 1 all-or-none process, can be used to dissect a complex learning process into elementary parts.

Since Rock's (1957) paper, a major issue in verbal learning has been whether paired associates are learned all-or-none or gradually. Beautiful examples of all-or-none learning have been erected and then smashed by iconoclasts with surprising frequency, and now the issue seems well matched—both sides claim victory.

It happens that no party to this controversy has seriously attempted to define the issue with any generality or detail. It has seemed intuitively clear what is meant by all-or-none learning, and the notion has mainly either been illustrated by experiment, as by Rock (1957), or embedded within ongoing learning theories (Estes, 1960; Restle, 1962).

The present paper will try to determine what it is that experiments like Rock's would prove, given they are valid. What *is* the experimental hypothesis of all-or-none learning? This paper will also attempt to determine what property a theory must have to be called an all-or-none theory. From this analysis it will become apparent that the idea of all-or-none learning supported by some experiments and the idea destroyed by other experiments may be quite different ideas.

It is unreasonable to assume that all learning, of all sorts and however defined, is all-or-none. It is uninteresting to assert that there exists some, perhaps very special, experimental procedure that yields all-or-none learning. The all-or-none hypothesis can be formulated with either a very broad or a very narrow *scope*, and of course the scope may also be discussed in more detail as to precisely which experiments will, and which will not, yield all-or-none data. Analysis of scope will show that the interesting experimental question is not "Is learning all-or-none?" whatever that means, but "What is the inner structure of a learning task, and how is that structure determined?"

ALL-OR-NONE DATA

Three different experimental paradigms have been put forward as tests of the all-or-none hypothesis.

1. Rock (1957) used paired-associates learning, beginning with a large pool of items. A control group merely learned a typical list in the usual way, reading the list of paired associates, then responding to the stimuli given alone. An experimental (replacement) group started with a similar list. The list of stimulus-response pairs was read, then a test (stimulus only) was given on the list. If any item was missed during the test, it was removed from the list and replaced with a fresh item. If the replacement group learns as rapidly as the control, this is evidence for all-or-none learning.

2. Estes (1960) proposed a miniature experiment, consisting of one training or rein-

---

[1] This research was supported by Grant GB 2848 from the National Science Foundation.

forced (R) trial of stimulus-response pairs, followed by two or more test (T) trials: the R-T-T experiment. If subjects who fail on T_1 perform no better than chance on T_2, this is taken as evidence of all-or-none learning.

3. Bower (1962) simply taught a list of items to a group of subjects, then compared the detailed statistics of the data with calculations from a mathematical model, the "one-element" model of stimulus-sampling theory. If the data agree with this model, it is said that all-or-none learning has been observed.

These experiments all deal with acquisition and with items which subjects master completely. Consider a sequence of training and test trials, all alike, and suppose the subject is to be scored as correct or wrong on each test. The all-or-none hypothesis, in each of the three experiments, asserts that unless the subject has completely learned the item, he has not learned it at all. Learning itself is not observable, but if the subject makes an error then he is known not to have learned. Rock's experiment discards the item, presumably throwing away any partial learning. Estes predicts that if an error has been made, performance on other tests will be at chance. Bower asserts that performance before any error is stationary, and that the future prospects following an error on Trial n are independent of n.

Since errors are the important indicators, an empirical theory of all-or-none learning may concentrate on errors. A concise and logically powerful definition comes from Feller's (1957) notion of a recurrent event. Let ϵ_n signify that an error occurs on Trial n, and c_n signify that a correct response occurs on n. Then define the probabilities:

$$\begin{aligned}
f_{n,1} &= P(\epsilon_{n+1} \mid \epsilon_n) \\
f_{n,2} &= P(\epsilon_{n+2} \text{ and } c_{n+1} \mid \epsilon_n) \\
&\cdots \\
f_{n,i} &= P(\epsilon_{n+i} \text{ and } c_{n+i-1} \text{ and} \cdots \text{and } c_{n+1} \mid \epsilon_n).
\end{aligned}$$

That is, $f_{n,i}$ is the probability that the next error is on $n + i$, given an error on Trial n. If $f_{n,i}$ is independent of n, then it is said that an error is a *recurrent event*. Each occurrence of an error resets the process to its origin. An error indicates that learning is incomplete, and since learning is all-or-none, if learning is incomplete there is no learning and the process is back at its origin. Mathematically, this definition had the advantage of referring only to an error and the next error, not to more complex configurations of data. However, given the fact that errors are recurrent events, and the set of probabilities $\{f_i\}$, it is possible to calculate the probabilities of all interesting events.[2]

Bower (1962) and Suppes and Ginsberg (1963) have insisted that in an all-or-none model performance before solution must be a sequence of Bernoulli trials. This would restrict the $\{f_i\}$ to the form $f_i = ab^i$, where a and b are probabilities.

This analysis of all-or-none learning is purely empirical and refers only to observables, namely, correct responses and errors on the successive trials. However, the definition is limited to acquisition of a simple problem and to an analysis into correct and wrong responses. If we want to know whether the underlying process of learning is all-or-none, we must go beyond simple acquisition experiments and provide tests different from the training conditions. To determine what a subject has learned, tests are made of stimulus generalization, memory, change of context, and change of response set (such as use of recognition versus recall tests of retention).

It may be argued that if learning is truly all-or-none, all these different tests should show all-or-none behavior. But then, is it necessary that there be no forgetting or that forgetting be all-or-none? It seems that a requirement of all-or-none forgetting may be different from the requirement of all-or-none learning. Similar uncertainties are involved in dealing with all other experimental procedures other than simple acquisition. More complete theory is needed to permit us to interpret more complicated and interesting tests of all-or-none learning.

---

[2] The argument for this point is based on the Estes and Suppes (1959) analysis of the foundations of learning models, in which the basic sample space is composed of sequences. Probabilities are assigned to events, which are compounded from cylinder sets of sequences. The probability of any set of outcomes on contiguous trials can be calculated from the probabilities $\{f_i\}$ in a simple way, and from this all interesting events can be constructed.

THEORIES OF ALL-OR-NONE LEARNING

Cognitive All-or-None Theory

Rock (1957) gave two accounts of all-or-none learning. In one, he characterized the results of his experiments by saying that repetition appeared to play no role. This formulation harks back to a kind of association theory in which contiguity, repetition, and other such factors produce associations. In conventional associationist theory, frequency plays the role of producing gradual increases of strength of the association. Rock's first statement that repetition plays no role merely says that learning is all-or-none.

In other discussions, Rock (1957), Rock and Heimer (1959), and Rock and Steinfeld (1963) assert a second theory that a paired-associate item is usually learned by forming a mnemonic device or "association." If the subject forms such an association he remembers the item well, and if he fails to form a suitable association he gains nothing and must try again next trial.

Rock's second theory is typically cognitive in that (*a*) learning involves active incorporation and recoding of the material, and (*b*) the true stimulus and response may not correspond in a simple way to the physical identification. The same nonsense syllable takes different psychological forms.

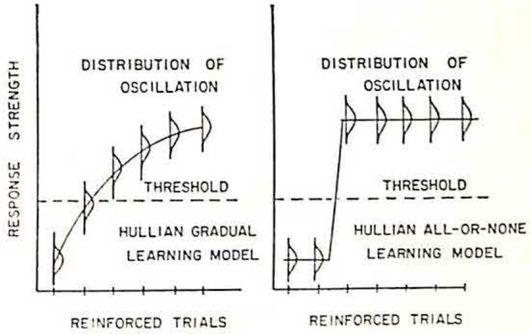


FIG. 2. Graphical illustrations of the usual Hullian gradual-learning model and a Hullian all-or-none learning model.

Rock's general theory has a mathematical form similar to cognitive theories of cue learning (Restle, 1962). Let H be the set of possible associations, mnemonic devices, or "strategies," corresponding to a particular pair of associates S-R. When shown S-R, the subject generates an association, that is, he samples from H, taking a Strategy h. If h is satisfactory, that is, if it enables the subject to remember the correct response, then it is presumably kept until the next test trial. If h does not mediate successful memory of the response, the subject resamples from the Universe H. If the subject resamples always from the same Universe H, then all-or-none learning must result, for after every failure to master the problem the subject returns to an initial state of sampling from H. This indicates that errors are recurrent events, hence the theory yields all-or-none data. The process is illustrated in Figure 1.

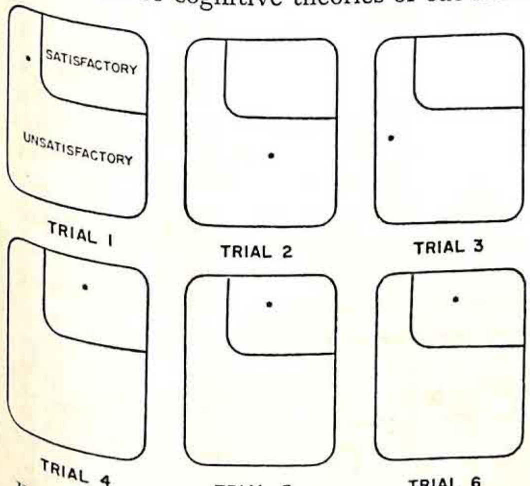


FIG. 1. Cognitive model of mastery of an item. (The first three tries yield unsatisfactory strategies and the subject resamples. On Trial 4 a satisfactory strategy is chosen and kept thereafter.)

Hullian All-or-None Theory

In Hullian theory the correct response will be made to a stimulus if its strength is above threshold and no other response is stronger. All measures of response result from the single response strength. Response strength oscillates randomly from trial to trial and ordinarily is incremented each trial of training.

To specify the all-or-none version of Hullian theory, merely assume that the correct response can take only two strengths, one below threshold and the other above (see Figure 2).

As habit strength shifts from subthreshold to maximum, the probability, latency, and amplitude of the response also must change from one value to a second, as must resistance to extinction and forgetting. Forgetting must

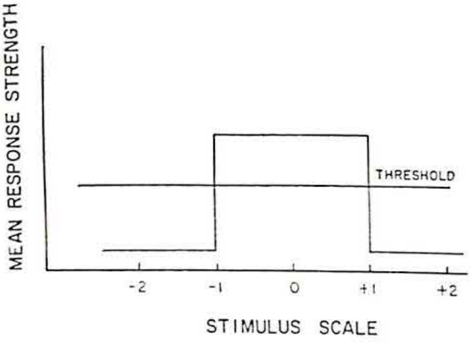


FIG. 3. All-or-none generalization gradient from Hullian all-or-none learning model.

be all-or-none for a given item and must be independent of overtraining on that item. If response strength takes only two values, then the "generalization gradient" cannot be gradual but must be as in Figure 3.

Furthermore, the various tests of response strength must all agree, except that they may set different thresholds. An example of the possibilities is shown in Figure 4, representing an easy recognition test, a recall test, and a difficult "quick recall" test. Notice that the subject gets correct responses before learning in the recognition test, not in either recall test; and the quick recall test is so difficult that the subject only gets about 60% correct after mastery.

This Hullian all-or-none theory is not satisfied by any known experimental procedures. It was put forth by Postman (1963) as an explication of the concept of all-or-none learning, along with considerable experimental disproof.

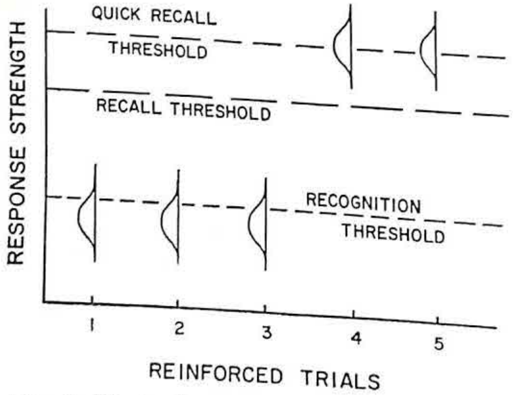


FIG. 4. Illustration of how different levels of performance can be obtained from Hullian all-or-none model by changes of the test conditions and resulting changes of response threshold.

Stimulus-Sampling All-or-None Theory

In Estes' stimulus-sampling theory, a stimulus is thought of as a set of stimulus elements. Each element is conditioned to exactly one response, and the probability of any response equals the proportion of elements conditioned to it. Learning consists of all-or-none conditioning of elements to responses.

All-or-none learning arises if the stimulus has only one element. In Bower's form of the theory, the element may be unconditioned (perhaps, conditioned to some inadmissable response) so that the subject guesses. The element becomes conditioned to the correct response on any training trial with Probability c (see Figure 5).

If stimuli are nonsense syllables or complex visual figures it is difficult to believe that each gives rise to a single element. Estes reconciles this observation with a "pattern theory," saying that the effective stimulus is a pattern, hence singular though complex. The pattern theory also explains how a subject may discriminate between stimuli with many common elements, how part-whole discriminations are performed, and why learning is retarded by stimulus variability and irrelevant variations of the stimuli.

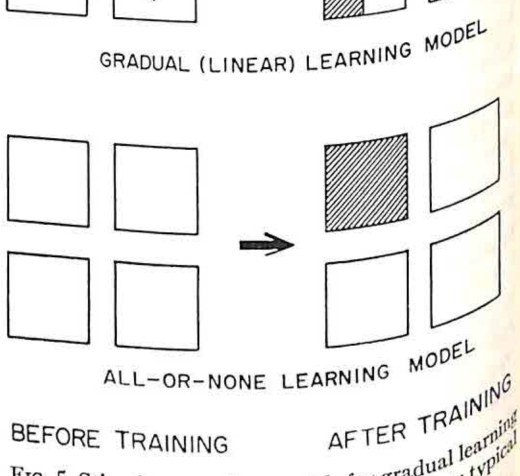


FIG. 5. Stimulus-sampling models for gradual learning (top) and all-or-none learning (bottom) of four typical items.

A difficulty of the simple pattern theory is that it treats each stimulus as unique, hence cannot make predictions of generalization or similarity. Friedman (1962) has discussed a "pattern plus components" theory in which the subject learns to respond to patterns and makes the learned response whenever the known pattern is presented. When a new stimulus pattern is presented, the subject has no response to the new pattern and uses the components (relatively numerous elements). It is assumed that if a pattern is conditioned to a response, its components also elicit the same response. This theory states that acquisition follows the all-or-none pattern theory and still accounts for graded stimulus generalization.

At present the writer has not seen a complete theoretical account of the interrelationships of patterns, subpatterns, and elements. An interesting situation, if it is possible, arises when a pattern is conditioned to one response and its elements to another. Now suppose that a stimulus situation very like the pattern is presented. If it is recognized as the same pattern, one response is made, but if it is taken to be a new pattern (although perhaps very similar) then response is by components and the other response is made. It is difficult to envisage the precise nature of this dividing line between another presentation of an old pattern and the presentation of a very similar new pattern.

Memory Models

If a paired-associate item is tested immediately after presentation it is almost always known (Peterson & Peterson, 1961). This simple fact has led to theories of paired-associate learning which suppose that the subject always learns, but then forgets before the test. Of course, if the subject just keeps forgetting between tests he never improves much over the first test and certainly cannot master the problem.

One approach is to suppose that the item is put in "temporary storage" from which it can be retrieved. In temporary storage the item can be erased or forgotten unless it is first transferred to "permanent storage." Friedman and Gelfand (1964) propose that the item is permanently stored if it is recalled successfully at a retraining trial. This theory would lead to very good memory after two immediately successive presentations, for then the subject would be very likely to remember from one to the other, and thereby fix the memory. Peterson, Saltzman, Hillner, and Land (1962) and Peterson, Hillner, and Saltzman (1962) have found instead that a relatively wide spacing of training trials is better, and Greeno (1964) has shown that there is almost no new learning on the second immediate presentation of an item. There seems to be a special virtue in remembering after a long pause—and the immediate memory model of Friedman and Gelfand does not handle this fact.

Theories with several levels of memory can lead to all-or-none learning or to two-stage models (Waugh & Smith, 1962). The major theoretical problem is to decide what levels of storage are possible and under what conditions an item transfers from one level to another.

The cognitive strategy-selection theory mentioned above is a memory model. When the subject selects a satisfactory strategy, the item is put in what memory theories call "permanent storage." Typically, memory strategies in paired-associates learning involve selection of a cue from the stimulus. If the cue selected is not unique to that stimulus, the subject will make generalization errors, then resample a new strategy. In memory terms, this says that the item was in temporary storage and was erased by a new item. If the subject's strategy is unsatisfactory for other reasons, memory theory might say that the subject has difficulty gaining access to the engram of the item, or that the memory dissipates. The strategy-selection theory has many different levels of storage, and these levels differ qualitatively as well as quantitatively, since a strategy may be unsatisfactory in many different ways.

GRADUAL LEARNING MODELS WHICH PRODUCE ALL-OR-NONE DATA

Several learning models postulate a transformation on response probabilities or underlying states or strengths. Any of several transformations is possible, depending upon the response of the subject and the outcomes of reinforcements. Within each of these general models, it is possible to set up a special "re-

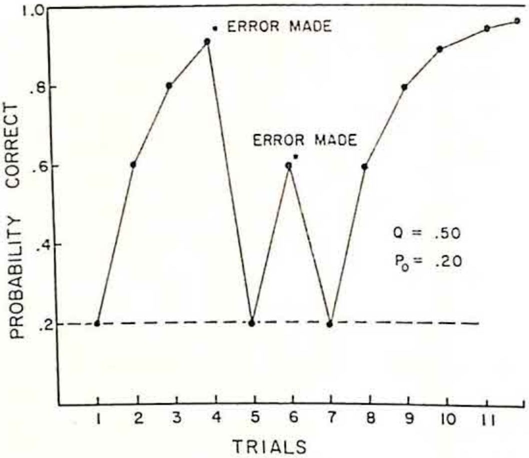


FIG. 6. Illustration of theoretical underlying probabilities of correct response for a subject who makes errors only on Trials 4 and 6. (Assumptions are those of a linear model in which the probability resets to P_0 at any error.)

current events model" satisfying our definition or all-or-none learning.

Linear All-or-None Model

The linear model (Bush & Mosteller, 1955) writes linear operators which operate on p, the probability of a certain response. In fixed-point form such operators are written:

$$Qp = (1 - \theta)p + \theta\lambda$$

where θ is learning rate and λ is asymptote. Let p be the probability of a correct response and the two operators be Q_E, applied when the subject makes an error, and Q_C, applied when he makes a correct response. To obtain an all-or-none model,

$$Q_C(p) = (1 - \theta)p + \theta$$
$$Q_E(p) = p_0.$$

A correct response increases p toward unity, but an error returns p to its initial value p_0 (See Figure 6).

The general linear model displays independence of path; given a level of p, nothing depends upon how it got there. Since errors set p at p_0, errors are recurrent events in this model. The f distribution is

$$f_i = (1 - \theta)^{i-1}(1 - p_0) \times \prod_{j=0}^{i-2} [1 - (1 - \theta)^j(1 - p_0)].$$

Since this is not a geometric sequence, the model predicts some "nonrandomness in the presolution period," if analyzed using Bower's methods. More specifically, runs of successes will tend to be somewhat longer than predicted by the Bower model. The probability of an infinite string of correct responses (what in an all-or-none model is taken for learning) is

$$f_\infty = \prod_{j=0}^{\infty} [1 - (1 - \theta)^j(1 - p_0)]$$

an infinite product. This function is positive for reasonable values of the parameters, such as, for example, $p_0 = .50$, $\theta = .40$. (See Bush & Mosteller, 1955, p. 160, who give a small table of values.)

From this model one can deduce, for example, that the total errors (**T**) will have the distribution

$$P(\mathbf{T} = k) = (1 - f_\infty)^k f_\infty$$

where f_∞ is as given above.

All-or-None Beta Model

Luce (1959) has proposed that response-strengths, nonnegative variables (v_i), be assigned to the alternative responses. In general, after a trial of type k, the new value of v_i for response i is given by

$$v_i(n + 1) = \beta_{ik}v_i(n)$$

for some $\beta_{ik} > 0$, depending on the response and outcome. For a simple example, suppose that v_E, the error-producing tendency, is fixed at 1. When a correct response is made, for some $\beta > 1$,

$$v_C(n + 1) = \beta v_C(n).$$

After an error,

$$v_C(n + 1) = v_0,$$

a constant. The error is now a recurrent event, since it returns the two v-scale values to their original values of 1 and v_0. The f distribution is

$$f_i = \frac{{v_0}^{i-1}\beta^{i(i-1)/2}}{\prod_{j=0}^{i} (1 + \beta^j v_0)}$$

and the probability of an infinite string of successes is

$$f_\infty = \prod_{i=0}^{\infty} \frac{v_0\beta^i}{(1+v_0\beta^i)}.$$

Similar results can be obtained with the general urn model of Audley and Jonkheere (1956), with the author's discrimination learning model (Restle, 1955), and almost all other continuous-learning models, by the simple expedients of (*a*) defining the responses as correct and wrong, and (*b*) resetting the model at its initial values whenever an error occurs.

All-or-None Hullian Model with Gradual Learning

Above, the Hullian notion of all-or-none learning was given, in which response strength takes only two values. It is also possible to obtain all-or-none data despite having gradual learning. One simple model works as follows: (*a*) oscillation has zero variance; (*b*) items have a uniform distribution from threshold to some fixed value m below threshold; and (*c*) during training each item climbs by a continuous exponential function from its initial value, x, to a value $m + x$. The model is shown in Figure 7. A fixed proportion (in Figure 7, one-half) of the previously below-threshold items cross threshold each trial. Presuming we have no knowledge of the preexperimental response strength of an item, the data will look exactly like all-or-none learning. This is a refinement of the interpretation of Underwood and Keppel (1962).

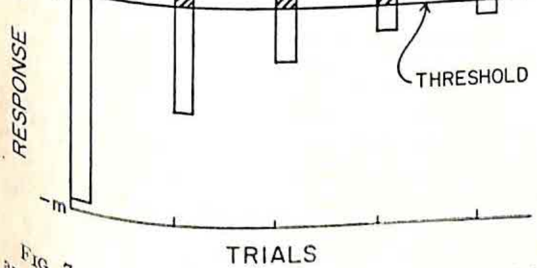


FIG. 7. A pseudo-Hullian model yielding data which appear to reflect all-or-none learning, although the actual underlying processes are of gradual learning. (In this case the spread of the distribution of response strengths is not changed by training.)

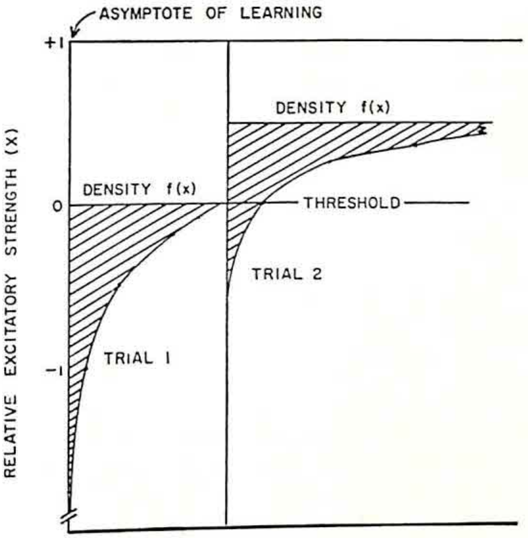


FIG. 8. A second Hullian model yielding data which appear to reflect all-or-none learning although the underlying processes are of gradual learning. (In this model response strengths all converge to a single asymptote with extended training.)

Another Hullian model has a single asymptote for all items, and a common learning rate. All items have zero oscillation. Assign value of 0 to threshold and value of 1 to the asymptote of response strength, and suppose that each item moves a fraction θ of the way toward asymptote from its initial value. The whole distribution is compressed by a factor of $1 - \theta$ each trial. To have all-or-none learning, a constant fraction must cross threshold each trial, and this happens if the cumulative distribution function is

$$\Phi(x) = e^{(1-e^{-\beta x})}$$

where $\beta = -\log_e(1-\theta)$ and x ranges from $-\infty$ to 0. Figure 8 gives a general idea of the shape of the original density function given by

$$\varphi(x) = \beta e^{\beta x} e^{1-e^{-\beta x}}.$$

This second Hullian model says that all items will be equally strong after extensive training, whereas the first model preserves the original distribution. Both produce perfectly all-or-none data at the phenotypical level, though both models are genotypically gradual learning models.

In summary, all-or-none learning can arise from versions of cognitive theory, Hullian theory, stimulus-sampling theory, and as a special case of linear learning models, Luce's beta model, and the usual Hullian gradual-learning model.

EXPERIMENTAL FINDINGS

The three main methods of experimenting are (*a*) the Rock replacement experiment and variations, (*b*) the Estes miniature experiment, and (*c*) ordinary learning with detailed analyses. Of these, the Rock technique, which involves a clever experimental innovation, has received the closest inspection. The results are as follows: Using the letters except I singly and doubly (AA, BB, etc.) as stimuli, and the numerals 1–50 as responses, with slow training (up to about 8 seconds per presentation, about 30 seconds between trials), everyone gets Rock's all-or-none result. Mastery of a list of 12 items takes four to five trials. Individual letter-number pairs differ widely in difficulty (Underwood, Rehula, & Keppel, 1962), so that the results may be affected by item selection—the more difficult items will more likely be missed, hence dropped from the experimental list and replaced by unselected items that on the average are easier. The quantitative effects of such selection are not yet worked out in detail for randomly selected lists but are under investigation (Postman, 1962; Williams, 1961, 1962).

Rock (1957) also used nonsense syllables as stimuli and responses, in his second experiment. Even with very slow training these are much more difficult: Rock's subjects, learning a list of eight pairs, took an average of over eight trials, and made about 3.2 errors per item. Rock and Heimer (1959), using nonsense syllables, found no difference between repeated (wrong) and new items, an indication of all-or-none learning. However, other experiments using about the same technique have not all obtained all-or-none results: Lockhead (1961) with slow training obtained all-or-none results, as did Bolles in a serial-learning task in which he scrambled or replaced items in the middle of the list after the list was partly learned. However, Battig (1962, Experiment 1) found a difference of 4.8 (experimental) versus 3.7 (control) errors per item, favoring repeated items. Brackett and Battig (1963, Experiment 1) found differences of 5.1 versus 2.7 errors per item. Kristofferson (1961) found differences of 11.4 versus 8.5 trials for a list of 15 CVCs, then repeated the experiment with only 12 pairs and got mean trials of 12.4 and 8.2. Lockhead (1961), with rapid training (0.75-second duration of items), found a very large difference in favor of repeated items. Postman (1962) found mean trials of 11.5 to 9.0, then a median of over 15 as compared with about 9 trials, in lists of eight items. Rock and Steinfeld (1963) found 8.6 versus 6.0 trials to criterion, with a list of eight items, and mean errors per item of 4.4–3.0, when subjects were required to spell the responses. Clark, Lansford, and Dallenbach (1960) repeated Rock's (1957) results and also showed that with rapid (1 second) presentation, learning was not all-or-none.

These results indicate: (*a*) Nonsense syllables are more difficult to learn than the letter-number pairs (at least twice as many trials to criterion, usually more). (*b*) The all-or-none finding, though repeatable with the letter-number pairs (and possibly influenced by item selection), is not consistently repeatable with CVC-CVC pairs.

Rock and Steinfeld (1963) note that spelling a CVC breaks it up into parts. Generally speaking, CVCs of an association value of about 50% are fairly but not extremely pronounceable. Under the best circumstances they are not highly organized, and learning the item may involve response integration. Furthermore, a list of eight CVC-CVC pairs contains 32 consonants and 16 vowels, requiring a reasonable amount of formal similarity. Restle (1964) found decided multiple-stage effects caused by stimulus discrimination between formally similar nonsense syllables used as stimuli, even with one-digit responses. Hence, learning of CVC-CVC pairs usually requires some response integration and some selective stimulus discrimination, and is likely to produce multiple-stage learning. The experimental data are on the borderline between all-or-none and gradual learning, using the Rock test.

Some other materials have been used, with varying results. Battig (1962), in a mixed-list design with a total list of 12 items, used nonsense forms said to be discriminable but with low association values. Learning was rapid (2.7–3.4 errors per item, about six trials), and results were not all-or-none. Kintsch (1963) used two-digit numbers as stimuli and CVCs as responses, and found two-stage learning. He used the valuable innovation of analyzing the control group's data by Bower's method, finding a two-stage model, then showing that the same model agreed with Rock's experimental results. Rock and Heimer (1959) used common first names as stimuli and responses in two studies, along with innovations of method including use of a matching test. Interestingly, though the responses are integrated and the stimuli presumably dissimilar, they found all-or-none learning only in one of three experiments. It may be that the obvious categories (especially boy versus girl) lead subjects to classify these materials and learn in a multistage manner.

In Estes' miniature experiment, the basic design is a training trial (R) followed by more than one test (T-T). The symptom of all-or-none learning is a high proportion correct on test two (T_2) given correct on T_1 and (more important) the probability of correct on T_2 given not correct on T_1; the probability $P(C_2|N_1)$ should be zero or chance.

These are essentially the results of Estes' report (Estes, 1961; Estes, Hopkins, & Crothers, 1960) in two R-T-T studies using CCC stimuli and common English words as responses, and in two free-recall studies using common English words. The probabilities of the four outcomes are shown in Table 1.

TABLE 1

PROPORTIONS OF OUTCOMES OBSERVED IN FOUR R-T-T STUDIES REPORTED BY ESTES (1960, 1961)

	Paired associates, CCC-word		Free recall of words	
	Exp. I	Exp. II	Exp. III	Exp. IV
$P(CC)$	.238	.35	.292	.151
$P(CN)$	.147	.14	.098	.129
$P(NC)$	.017	.05	.003	.014
$P(NN)$	.598	.46	.606	.705

The probability is supposed to fall mainly in three cells: CC (the subject learned), NN (the subject did not learn), and CN (the subject learned but forgot). In the paired-associates learning, some guessing also may occur.

Jones (1962) repeated the above studies using a consonant as stimulus and a numeral, 1–8, as response. The results were:

$$P(CC) = .398$$
$$P(CN) = .133$$
$$P(NC) = .089$$
$$P(NN) = .380$$

which generally agrees with Estes' result but shows faster learning and a high frequency of NC events. Jones delayed the second test.

Jones also did an R-T-T-T-T experiment, saying that if a subject is wrong on T_1 he must not have learned, hence must do badly throughout the tests (if the all-or-none principle is to hold). Let C_i mean a correct response on Test Trial i, and N_i mean an error on Trial i. The all-or-none assumption implies (with $i > 1$) that

$$P(C_i|N_1) = k$$

where k is some constant guessing probability. A severe test of this hypothesis is to consider items missed on the first test but responded to correctly on some other tests; by theory, if the subject guesses wrong at the first test and correctly on Tests 2 and 3, he still has only a chance probability of getting a correct answer at Test 4. Jones found that this theoretical expectation failed:

$$P(C_2|N_1) = .29$$
$$P(C_3|N_1, C_2) = .61$$
$$P(C_4|N_1, C_2, C_3) = .77.$$

There are two possible explanations of this finding. One is that, contrary to the simple all-or-none theory, the items do differ in probability of correct response. Hence, most items with the response profile $N_1C_2C_3$ have a relatively high response probability, whereas those with a known profile of only N_1 have a lower probability. A second hypothesis is that learning occurs during test. It might be that whenever the subject responds correctly during test he reinforces the correct response, and therefore possibly might shift from the guessing to the learned state during the C_2 and C_3 trials.

This second hypothesis is compatible with the idea that learning is all or none. If the result was due to different levels of probability, the patterns N, NC, and NCC merely single out different subsets of subjects and therefore produce the different conditional probabilities observed. The calculation can be run backwards with precisely the same effect; that is, one can single out items in which the last (rather than the first) response is wrong, and then make the parallel conditional probabilities. The result of this calculation, carried out from Jones' data, is as follows:

$$P(C_3|N_4) = .25$$
$$P(C_2|N_4, C_3) = .45$$
$$P(C_1|N_4, C_3, C_2) = .67.$$

This shift in probabilities cannot, of course, be attributed to learning since it progresses in the wrong direction. Postman (1963) and Schwartz (1963) have done more complex experiments involving both several reinforced trials and several tests and also have obtained results not in accord with the simple all-or-none hypothesis.

The Rock experiment tests whether one failure equals zero training, and the R-T-T experiment deals with a single training trial. Following techniques of Bush and Mosteller (1959), Bower (1961) worked out many consequences of an all-or-none model and then compared theory and experimental results in great detail. One study used 10 consonant pairs as stimuli, the numerals 1 and 2 as responses, and self-paced training by flash card. Verbal discrimination (Bower, 1962) has yielded similar results. Numerous studies, many unpublished, have found that with discrete stimuli and two simple responses, all-or-none data can readily be obtained.[3]

Using Restle's (1962) model, Trabasso (1961, 1963) demonstrated all-or-none learning in concept formation, and Bower and Trabasso (1964) have extended the results using a variety of experimental methods including several related to Rock's technique.

COMPOUNDS OF ALL-OR-NONE LEARNING

One alternative to the all-or-none hypothesis is that learning proceeds in stages, and the stages are all-or-none. Atkinson and Crothers (1963) reported eight experiments, in all of which a two-stage all-or-none model gave a good fit to the data. In the model, the states are S_0 (prelearning), S_1 (an intermediate level of performance), and S_2 (solution). The transitions are shown in Figure 9.

Kintsch (1963) employed a similar model to predict the outcome of a Rock-type substitution experiment. Waugh and Smith (1962) analyzed free verbal recall with several processes—labeling, processing, and storing—in a model like that in Figure 9.

Restle (1964), in a long theoretical article, discussed several experiments, in particular one by Bower and Levine (unpublished) and another by Restle and Trabasso. In the Bower-Levine experiment, each stimulus was three Korean characters. Nine stimuli, constructed in three subsets of three stimuli each, were given in a list. In one subset, the three stimuli had no elements in common. In another subset, the three stimuli had one Korean character in common. In a third subset, the three stimuli had two characters in common. Restle's analysis revealed all-or-none learning for the stimuli with no common elements, but distinct two-stage learning for items with overlap.

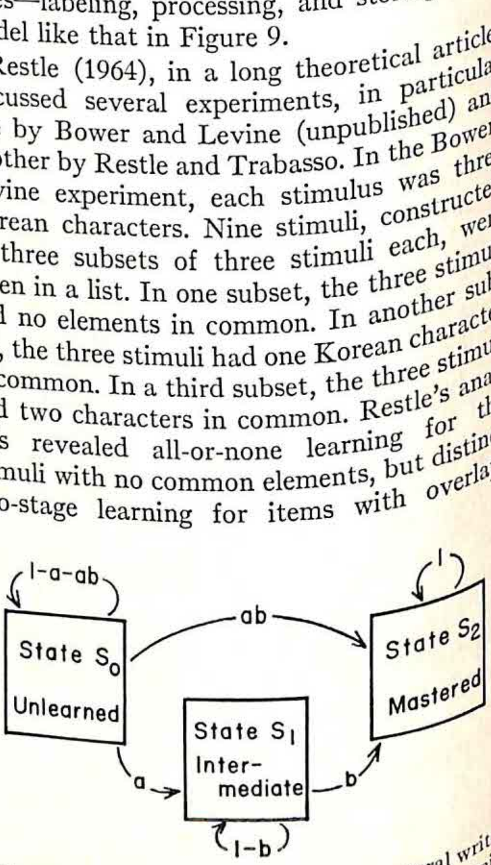


FIG. 9. Two-stage model employed by several writers to account for data in which simple all-or-none learning is not exhibited.

---

[3] The writer has devised a classroom demonstration of all-or-none learning. Stimuli, read orally to a class, are two-syllable paralogs such as JEDIN, NABOG, etc. Responses are check or plus. Very good fits to the Bower model are obtained. However, recently one subject remarked in class that his method had been to learn to recognize all the words going with plus. All others would be called check. Thus this two-choice paired-associates problem was learned as a simple recognition task. Approximately 200 of 230 students in the class indicated that they had used similar strategies. This means that two-choice paired-associates learning may not involve any association learning at all.

The second stage was identified as discrimination learning, and it was ascertained that the excess errors were stimulus-generalization errors. The model is that shown in Figure 9.

Restle and Trabasso's experiment used nonsense syllables, some paired as formally similar (VEH, VUH) and others unique. Responses were numbers 1–8 or 1–4. Detailed analysis showed strong stimulus generalization, some unintended stimulus generalization between syllables with the same vowel (PIW, YIG), apparent response generalization, and, during guessing, strong response preferences. The conclusion was that one-, two-, three-, or even four-stage learning could be picked apart and demonstrated. (See also McGuire, 1961.)

Polson, Restle, and Polson (1965) used pictures as stimuli and common nouns as responses. Some pairs of pictures in the list of 16 were twinned: 2 oak leaves, 2 quadrilaterals of similar shape, etc. Other stimuli were unique in the list. Items with unique stimuli were learned in almost all-or-none fashion, whereas items with twinned stimuli were learned in two stages, and the second stage was identified as stimulus discrimination. Since it was assumed that stimulus discrimination depended on selection of a differential cue, an event independent of forming the original association, the model derived was that of Figure 9.

Crothers (1964) taught subjects compound responses in paired associates; responses were two-dimensional, like "red square" or "green triangle." He showed that mastery of such responses was a compound of mastery of the parts. Hillner (1965) taught paired associates in which the responses were strings of three unrelated common words, (BANK-YARN-MILK). If the second and third words were each introduced only after the previous segment of the response was mastered, the data corresponded closely with an extention of the model of Figure 9, but if the whole compound response was introduced at once, the data revealed a strong correlation between errors made on the various components of the response.

In all of these studies, the multiple-stage model was evaluated by its ability to fit the data, but the assumptions of the model were also evaluated by traditional experimental methods of manipulating experimental variables and classifying the types of errors observed. Conclusions were properly based only on the quantitative agreement of converging, independent operations.

DISCUSSION

Rock emphasized the importance of mnemonic devices among which the subject would sample in solving the problem of remembering. Estes employed the Gestalt idea that the stimulus would act as a united pattern. Restle reintroduced the idea of selection of hypotheses. These theories are usually set in opposition to pure associationist theory in which learning is a more passive process resulting from conjunction of stimulus, response, and perhaps reinforcement, in which the typical stimulus is analyzed into small elements and in which learning is the formation of simple habits rather than the selection among more complex strategies.

All continuous-learning (associationist) theories can lead to all-or-none data, and the various cognitive theories of all-or-none learning differ among themselves in important respects. Rock's mnemonic devices may lead to associative connection rather than selection of hypotheses and do not require that the stimulus be a pattern. Estes' pattern theory can be a standard passive-association theory, requires no complicated strategies, and is more association than selection. Restle's strategy-selection theory makes no essential use of mnemonic devices, and does not assume that the stimulus is organized.

Since several different association theories, and several different cognitive theories, all have the capacity to explain all-or-none learning, the simple fact of all-or-none learning is not appropriate for the decisive confrontation of major theoretical approaches. No classical theoretical issue can be decided by answering whether learning is all or none.

Logically, there are several versions of the all-or-none hypothesis, including: (*a*) All learning is all-or-none; (*b*) There exists at least one task that is learned all-or-none; (*c*) A single item of a paired-associate list is learned all-or-none; (*d*) A single item of a paired-associate list, provided the list involves no stimulus confusion, response integration, etc., may be learned all-or-none; (*e*) Concept formation, discrimination learning, and certain other tasks are learned all-or-none; (*f*) Any learning task that involves only one difficulty is learned all-or-none.

Of these, *a* is certainly false, and *c* and *e* are no doubt insufficiently qualified. Version *b*, the weakest, is true at least to an excellent approximation. Version *d* is the one most often intended in the experimental literature, and the experimental debate centers around whether Rock's experiment, or Estes', actually is learned all-or-none, and whether these and other experimenters have succeeded in removing all but one source of difficulty, usually the "association phase." Recent work by Trabasso (1961, 1963) and Bower and Trabasso (1964) supports Version *e*, with certain restrictions that are not yet well specified.

The strongest possible interpretation of all-or-none learning is Version *f*. This is not merely an assertion that some example of all-or-none learning exists, but is a general hypothesis that attempts to specify the necessary and sufficient conditions for all-or-none learning. Unfortunately for the experimenter, it is not known a priori exactly what must be done to remove all but one source of difficulty. If, in some experiment, all-or-none learning is not obtained, it would be possible to respond that the problem must still involve more than one source of difficulty. This shifts the burden of proof to the proponent of all-or-none learning, who must give independent evidence for the existence of several sources of difficulty.

The advantage of Version *f* is that it permits an account, within the all-or-none framework, of data that do not show the statistical all-or-none properties. If a given problem presents the subject with several difficulties, then its solution can be described by a multiple-stage all-or-none model.[4] Such models encompass the complexities usually studied in the human-learning laboratory, yet have a manageable mathematical form. The task ahead is to extend the precise logic of all-or-none models to the analysis of more complex and interesting tasks.

---

[4] The phrase all-or-none denotes the basic assumption of a Markov chain; therefore, these proposals are not at all new, but now have relatively solid empirical support (see Miller, 1952).

REFERENCES

Atkinson, R. C., & Crothers, E. J. Tests of acquisition and retention axioms for paired-associate learning. Technical Report No. 60, October 25, 1963. Institute for Mathematical Studies in the Social Sciences, Stanford University.

Audley, R. J., & Jonckheere, A. R. The statistical analysis of the learning process. II. Stochastic processes and learning behavior. *British Journal of Statistical Psychology*, 1956, **9**, 87–93.

Battig, W. F. Paired-associate learning under simultaneous repetition and nonrepetition conditions. *Journal of Experimental Psychology*, 1962, **64**, 87–93.

Bower, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, **26**, 255–280.

Bower, G. H. An association model for response and training variables in paired-associate learning. *Psychological Review*, 1962, **69**, 34–53.

Bower, G. H., & Trabasso, T. R. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1964.

Brackett, H. R., & Battig, W. F. Method of pretraining and knowledge of results in paired-associate learning under conditions of repetition and nonrepetition. *American Journal of Psychology*, 1963, **76**, 66–73.

Bush, R. R., & Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.

Bush, R. R., & Mosteller, F. A comparison of eight models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959.

Clark, L. L., Lansford, T. G., & Dallenbach, K. M. Repetition and associative learning. *American Journal of Psychology*, 1960, **73**, 22–40.

Crothers, E. J. All-or-none paired-associate learning with compound responses. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1964.

Estes, W. K. Learning theory and the new "mental chemistry." *Psychological Review*, 1960, **67**, 207–223.

Estes, W. K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika*, 1961, **26**, 73–84.

Estes, W. K., Hopkins, B. L., & Crothers, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *Journal of Experimental Psychology*, 1960, **60**, 329–339.

Estes, W. K., & Suppes, P. Foundations of linear models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959.

Feller, W. *An introduction to probability theory and its applications*. (2nd Ed.) New York: Wiley, 1957.

Friedman, M. P. Application of statistical learning theory to component-vs.-compound discrimination learning. Unpublished doctoral dissertation, Indiana University, 1962.

Friedman, M. P., & Gelfand, H. Transfer effects in discrimination learning. *Journal of Mathematical Psychology*, 1964, **1**, 204–214.

GREENO, J. G. Paired-associates learning with massed and distributed repetitions of items. *Journal of Experimental Psychology*, 1964, **67**, 286–295.

HILLNER, K. P. The response learning stage and the mathematical analysis of paired-associate learning. Unpublished doctoral dissertation, Indiana University, 1965.

JONES, JOAN E. All-or-none versus incremental learning. *Psychological Review*, 1962, **69**, 156–160.

KINTSCH, W. All-or-none learning and the role of repetition in paired-associate learning. *Science*, 1963, **140**, 310–312.

KRISTOFFERSON, A. B. Repetition and paired-associates learning. *Science*, 1961, **134**, 2036–2037.

LOCKHEAD, G. R. A re-evaluation of evidence for one-trial associative learning. *American Journal of Psychology*, 1961, **74**, 590–595.

LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.

MCGUIRE, W. J. A multiprocess model for paired-associate learning. *Journal of Experimental Psychology*, 1961, **62**, 335–347.

MILLER, G. A. Finite Markov processes in psychology, *Psychometrika*, 1952, **17**, 149–167.

PETERSON, L. R., HILLNER, K., & SALTZMAN, D. Supplementary report: Time between pairings and short-term retention. *Journal of Experimental Psychology*, 1962, **64**, 550–551.

PETERSON, L. R., & PETERSON, M. J. Minimal paired-associate learning. *Journal of Experimental Psychology*, 1961, **63**, 521–527.

PETERSON, L. R., SALTZMAN, D., HILLNER, K., & LAND, V. Recency and frequency in paired-associate learning. *Journal of Experimental Psychology*, 1962, **63**, 396–403.

POLSON, M., RESTLE, F., & POLSON, P. Association and discrimination in paired-associates learning. *Journal of Experimental Psychology*, 1965, **69**, 47–55.

POSTMAN, L. Repetition and paired-associate learning. *American Journal of Psychology*, 1962, **75**, 372–389.

POSTMAN, L. One-trial learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning*. New York: McGraw-Hill, 1963.

RESTLE, F. A theory of discrimination learning. *Psychological Review*, 1955, **62**, 11–19.

RESTLE, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 329–343.

RESTLE, F. Sources of difficulty in learning paired associates. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1964.

ROCK, I. The role of repetition in associative learning. *American Journal of Psychology*, 1957, **70**, 186–193.

ROCK, I., & HEIMER, W. Further evidence of one-trial associative learning. *American Journal of Psychology*, 1959, **72**, 1–16.

ROCK, I., & STEINFELD, G. Methodological questions in the study of one-trial learning. *Science*, 1963, **140**, 822–824.

SCHWARTZ, M. Transfer from failed pairs as a test of one-trial vs. incremental learning. *American Journal of Psychology*, 1963, **76**, 226–273.

SUPPES, P., & GINSBERG, R. A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning, with application to concept formation in children. *Psychological Review*, 1963, **70**, 139–161.

TRABASSO, T. R. The effect of stimulus emphasis on strategy selection in the acquisition and transfer of concepts. Unpublished doctoral dissertation, Michigan State University, 1961.

TRABASSO, T. R. Stimulus emphasis and all-or-none learning in concept identification. *Journal of Experimental Psychology*, 1963, **65**, 398–406.

UNDERWOOD, B. J., & KEPPEL, G. One-trial learning? *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 1–13.

UNDERWOOD, B. J., REHULA, R., & KEPPEL, G. Item selection in paired-associate learning. *American Journal of Psychology*, 1962, **75**, 353–371.

WAUGH, N. C., & SMITH, J. E. K. A stochastic model for free recall. *Psychometrika*, 1962, **27**, 141–154.

WILLIAMS, J. P., Supplementary report: A selection artifact in Rock's study of the role of repetition. *Journal of Experimental Psychology*, 1961, **62**, 627–628.

WILLIAMS, J. P. A test of all-or-none hypotheses for verbal learning. *Journal of Experimental Psychology*, 1962, **64**, 158–165.

(Received April 12, 1965)

Psychological Bulletin
1965, Vol. 64, No. 5, 326–329

CONTROL OF SPURIOUS ASSOCIATION AND THE RELIABILITY OF THE CONTROLLED VARIABLE

DANIEL KAHNEMAN[1]

Hebrew University, Jerusalem

The techniques of matched groups, analysis of covariance, and partial correlation represent various approaches to the prevention of a spurious association between X_1 and X_2 due to a confounding variable, X_3. In all these techniques the use of an unreliable measure for X_3 leads to a systematic bias of undercorrection. Adequate corrections are possible for the case of known reliability of X_3. Groups should be matched on true scores rather than observed scores, but no correction is possible for the factorial design in which groups are formed on the basis of unreliable correlated measures. Partial correlations should be corrected for the effects of unreliability of the controlled variable. Spuriously high partials are usually obtained when this correction is not applied.

Spurious correlations and confounding variables present a characteristic and recurrent problem to the social scientist. When samples of children drawn from several social classes (Variable X_1) are found to differ on the level of their professional aspirations (X_2), one asks: Are the differences due, at least in part, to social class differences in intelligence (X_3)? If a positive correlation is discovered between occupational aspirations (X_1) and college grades (X_2), the possibility that differences in attitudes toward achievement (X_3) mediate the relationship is immediately suggested.

The techniques most commonly used to remove confounding effects of X_3 on the relationship between X_1 and X_2 are various matched-group designs, the analysis of covariance, and partial correlation. Having used these techniques, the investigator is likely to feel safe: the confounding variable has been removed and need not be considered further in the interpretation of results.

The purpose of the present note is to draw attention to the effects of unreliability in the measurement of X_3 on these research designs. Such unreliability typically leads to undercorrection so that the spurious effects due to X_3 are only partly removed. Some procedures which fully correct these biases will be mentioned for the special case of known reliability of the controlled measure. No solutions are proposed to the difficulties of significance testing which arise because of the reliability issue.

Matched-Group Designs and the Analysis of Covariance

The essence of the matched-group design is that cases are selected from two or more populations so that known differences among these populations in the distribution of some variable are ostensibly removed. Thus, two groups of children drawn respectively from an upper- and lower-class background may be equated on the basis of intelligence test scores. Such a matching procedure may be followed in an attempt to study the effect of social class on some dependent variable without contamination by intelligence. Thorndike (1942) has pointed out that such an approach tends to lead to spurious results when the variable which is the basis for matching is measured with less than perfect reliability. The core of the fallacy is that groups matched in this manner are similar in their observed scores but differ in their *true* scores. On retesting, the lower-class children will regress toward the mean of the population from which they have been drawn and will tend to do more poorly. The middle-class children will regress toward their population mean, and their test scores will indicate improvement.

The magnitude of the regression effect can be calculated by Formula 1, if the reliability

---

[1] My thanks are due to Ester Samuel, of the Department of Statistics, for her help in viewing the problem in its proper perspective. J. Levin and M. Kubovy read the manuscript and made many helpful comments.

of the test within each population is known.

$$\bar{T} = \mu + \rho(\bar{X} - \mu), \qquad [1]$$

where $\bar{T}$ and $\bar{X}$ are, respectively, the means of true and observed scores in the group; ρ is the reliability coefficient, and μ is the mean observed score in the population.

The between-group differences in true scores of X_3 which remain after matching on observed scores make it likely that a spurious association between X_1 and X_2 will be found. The matching procedure undercorrects for such effects. Thorndike (1942) proposed that groups be matched on the basis of predicted true scores, recently termed "regressed scores" by McNemar (1962, p. 161). Formula 1 may be used for this purpose.

The caution suggested by Thorndike appears to have been largely ignored in the psychological literature, and the procedure of matching on observed scores is still commonly used. The fallacies due to unreliability and regression effects are particularly severe in a complex variant of the matched-groups design, the factorial design where the levels on the various factors are formed by selection of groups on the basis of test scores. Regression fallacies occur when measures of low reliability which are correlated in nature are made ostensibly orthogonal in the experimental population. This design is quite popular in personality research and has often been applied to measures of doubtful reliability.

A hypothetical example of this type of design is presented in Table 1. It is assumed that an investigator sets up a 2 × 2 factorial design on the basis of test scores on two measures, A and B. The test-retest reliabilities of these measures are .49 and .64, respectively. The correlation between them in an unselected population is .30. The table shows the mean true values to be expected on each variable when the mean observed standard scores for the high and for the low groups are .50 and −.50.

The values presented in Table 1 have been computed by multiple-regression equations, in which the observed values of A and B are used to predict the true scores on these variables. All scores are in standard units. The subscripts *o* and *T* refer to observed and true scores, respectively. The correlations used in the computation are: $r_{A_0B_0} = .30$, $r_{A_0A_T} = .70$, $r_{B_0B_T} = .80$, $r_{A_0B_T} = .43$, $r_{A_TB_0} = .38$. The correlations between true and observed scores are obtained by means of the appropriate corrections for attenuation. It is apparent in Table 1 that the true scores on the variables are neither matched nor orthogonal. Spurious main effects and interactions are a very likely outcome when this design is used with correlated measures of low reliability.

TABLE 1

MEAN TRUE SCORES OF GROUPS IN A FACTORIAL DESIGN FOUNDED ON OBSERVED SCORES

	Low, $\bar{A}_o = -.50$	High, $\bar{A}_o = .50$
High, $\bar{B}_o = .50$	$\bar{A}_T = -.19$ $\bar{B}_T = +.30$	$\bar{A}_T = +.43$ $\bar{B}_T = +.45$
Low, $\bar{B}_o = -.50$	$\bar{A}_T = -.43$ $\bar{B}_T = -.45$	$\bar{A}_T = +.19$ $\bar{B}_T = -.30$

It is important to note that this type of flaw is actually fatal to the design. There is no correcting formula which will simultaneously achieve the required matching of values in the rows and columns.

The analysis of covariance has been suggested as an alternative to the matching procedure which presumably overcomes the regression fallacy (McNemar, 1962, p. 373). The covariance design is generally used when there exists a substantial correlation between the dependent (X_2) and controlled (X_3) variables. Two main cases may be distinguished (Edwards, 1950) depending on whether systematic between-group differences exist on the controlled variable. In the absence of large between-group differences on X_3, unreliability in the measurement of this variable simply reduces the effectiveness of the analysis without producing any systematic bias. The case where large differences in X_3 are found is the one where matching and analysis of covariance are reasonable alternatives. Under such conditions, it can be shown that unreliability in X_3 has similar effects on the two procedures.

Consider the case where the null hypothesis is valid. When the reliability of X_3 decreases, the variance of this measure is inflated by error. Therefore, the regression of individual scores of X_2 on the observed values of X_3 within groups is of lesser slope than the corresponding regression on true values of X_3. However, when real differences exist among the means of X_3 in the different populations (classes of X_1), unreliability of measurement has relatively less effect on the between-group variance of X_3: the variance of observed means is partly due to sampling error (which is affected by reliability), but it is also affected by population differences in X_3, on which reliability has no effect. Consequently, the observed regression among group means will tend to be steeper than the within-group regression, which leads to a bias in the F test. It is interesting to note that increases in sample size have no systematic effect on the expected value of within-group estimates, while such increases consistently improve the approximation of between-group estimates to the true regression. Thus, in the case described, there is an inflated probability of Type I errors over stated significance levels which is due to insufficient control of X_3. The probability of such errors increases as the reliability of X_3 decreases; it also increases with sample size.

The conclusion must be that the analysis of covariance is suspect of undercorrection whenever prior analysis suggests that real differences among groups exist on the X_3 variable, unless this variable is identified with very high reliability.

Partial Correlation

Partial correlations are used: (*a*) to establish whether an observed correlation between two variables is spurious, being due to the joint association of these two variables with a third, or (*b*) less frequently, to uncover a correlation between X_1 and X_2 which may be obscured when these variables are differently related to X_3. In both cases it can be shown that unreliability in the measurement of X_3 biases results in a systematic and predictable manner.

The partial correlation between observed values of X_1 and X_2—with *true* scores in X_3 controlled—is given in Equation 2:[2]

$$R_{12.3} = \frac{r_{12} - R_{13}R_{23}}{\sqrt{(1 - R_{13}^2)}\sqrt{(1 - R_{23}^2)}}. \quad [2]$$

When X_3 is measured with reliability ρ, coefficients which involve this variable are attenuated. The partial correlation, with observed scores of X_3 controlled, now becomes:

$$r_{12.3} = \frac{r_{12} - \rho R_{13}R_{23}}{\sqrt{(1 - \rho R_{13}^2)}\sqrt{(1 - \rho R_{23}^2)}}. \quad [3]$$

When reliability is perfect, it can be seen that $r_{12.3}$ is equal to $R_{12.3}$. With zero reliability, $r_{12.3}$ is equal to r_{12}, that is, the measure of direct association which the use of partial correlation was intended to correct.

With intermediate values of reliability, the observed partial generally takes a value between the true partial and the direct correlation between X_1 and X_2.[3] Where the true partial is higher than the direct correlation, $r_{12.3}$ underestimates the relationship. Perhaps more important, when the true partial is substantially lower than r_{12}, the observed partial systematically overestimates the degree of unconfounded association between X_1 and X_2. In either case, the use of observed scores for X_3 in the partial coefficient yields a result which is biased in the direction of r_{12}. The covariance between X_1 and X_2 which is due to X_3 is not effectively partialled out.

The magnitude of this bias is illustrated in the following example. Consider the following correlations, in which true scores of X_3 are assumed: $r_{12} = .30$, $R_{13} = .60$, $R_{23} = .50$. The true partial correlation is 0. When X_3 is measured with a reliability of .64, the following observed correlations are expected: $r_{12} = .30$, $r_{13} = .48$, $r_{23} = .40$. The observed partial correlation is .135.

In order to obtain unbiased estimates of a partial correlation in practice, it is essential to correct r_{13} and r_{23} for attenuation due to the unreliability of X_3. Unlike the case of simple correlation, the correction for attenuation does not necessarily yield a higher estimate of the partial correlation. In the most common case, where the relationship between

---

[2] The symbol R is used for coefficients in which true values of X_3 are considered.

[3] This statement is not always true. Under some restricted conditions, the observed partial does not vary monotonically when the reliability of X_3 increases. The conditions for this effect are closely similar to the conditions under which $R_{12.3}$ and r_{12} are identical. In such cases, partial correlation is of no practical relevance. Its value is then affected to a very limited extent by the reliability of X_3.

X_1 and X_2 is suspected of being spurious, correction for attenuation will yield a lower value for the partial correlation. In general, high values of the observed partial will be found less biased than lower values. However, when the observed partial is of borderline significance, considerations of reliability may have an important effect on research conclusions.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

MCNEMAR, Q. *Psychological statistics* (2nd ed.). New York: Wiley, 1962.

THORNDIKE, R. L. Regression fallacies in the matched groups experiment. *Psychometrika*, 1942, **7**, 85–102.

(Received August 16, 1964)

Psychological Bulletin
1965, Vol. 64, No. 5, 330–338

ON THE PARTITIONING OF CONTINGENCY TABLES[1]

N. JOHN CASTELLAN, Jr.[2]

Institute of Behavioral Science, University of Colorado

The chi-square test is presented in general. If a contingency table is greater than 2 × 2 the usual chi-square test of independence fails to determine the source of interaction if interaction is present. Methods for extending the chi-square analysis of r × c contingency tables are presented. The procedures outlined consist of partitioning an r × c contingency table into (r − 1)(c − 1) independent components, each with 1 degree of freedom. A special partition of a 2 × c table into 3 independent components is also given. The results of partitioning permit precise inference about independence or interaction in a contingency table and allow specific hypotheses to be tested without loss of generality.

Most psychologists are acquainted with the use of the chi-square test to test a null hypothesis of independence of proportions in a fourfold contingency table and as a simple goodness-of-fit test. Behavioral science statistics texts generally give the chi-square test little attention beyond these two applications. However, the chi-square test has much appeal because of its apparent simplicity, for it is a test which requires only the "counting" of observations and the subsequent placing of observations into mutually exclusive and exhaustive categories.

Much more sophisticated use of the chi-square test can be made than is indicated in current behavioral science statistics texts. Notable discussions extending the application of the chi-square test are found in Cochran (1954), Wilson (1956), and Maxwell (1961). The purpose of this paper is to discuss a particularly useful application of the chi-square test which provides better opportunities for inference than is possible with the familiar simple test of independence of porportions. The procedures to be outlined allow efficient use of the degrees of freedom when testing for independence and allow more precise inference about independence.

In the sequel it is assumed that the reader is acquainted with the use of the chi-square test. No derivations are presented, but reference to discussions of important proofs will be made.

Chi-Square Test

If we have a bivariate (joint) distribution of two discrete random variables or sets of categories Y and Z, then a bivariate frequency distribution of a sample of N observations on (Y, Z) may be constructed. Such tables are usually called contingency tables and for convenience we will consider the table to have r rows and c columns, that is, the number of values or categories on Y and Z are r and c, respectively. If we record observed frequencies in the cells of such a table, then we obtain a table (see Table 1) where a_{ij} is the frequency of occurrence of the pair (Y_i, Z_j). We may sum the frequencies in rows and columns and obtain the marginal frequencies, that is, the frequencies of the values of Y and Z taken individually rather than jointly:

$$a_{i.} = \sum_{j=1}^{c} a_{ij}, \quad i = 1, 2, \cdots, r, \qquad [1]$$

$$a_{.j} = \sum_{i=1}^{r} a_{ij}, \quad j = 1, 2, \cdots, c, \quad \text{and} \qquad [2]$$

$$N = \sum_{i=1}^{r}\sum_{j=1}^{c} a_{ij}. \qquad [3]$$

Now, the contingency table may be viewed as a set of observations—that is, all distinct cells of the table correspond to distinct observations. Suppose we wish to compare the observed frequencies with some theoretical frequencies, Np_{ij}, where p_{ij} is the probability of an observation (Y_i, Z_j), that is, the probability of making an observation which would fall in the ijth cell of the contingency table, and N is the total sample size as defined by Equation 3 above. As with the observed frequencies we may obtain marginal probabilities $p_{i.}$ and $p_{.j}$, where

---

[1] Preparation of this paper was supported in part by the Tri-Ethnic Research Project, Grant No. 3M-9156 from the National Institute of Mental Health to the Institute of Behavioral Science, University of Colorado, O. C. Stewart and R. Jessor, Co-Directors. This paper comprises publication No. 45 of the Institute. The author would like to thank D. E. Bailey and S. W. Link for their helpful comments and cogent criticisms.

[2] Now at Indiana University.

$$p_{i.} = \sum_{j=1}^{c} p_{ij}, \quad i = 1, 2, \cdots, r, \quad \text{and} \qquad [4]$$

$$p_{.j} = \sum_{i=1}^{r} p_{ij}, \quad j = 1, 2, \cdots, c. \qquad [5]$$

Then it can be shown, regardless of the probability distribution of (Y_i, Z_j), that as N becomes large $(N \rightarrow \infty)$, the quantity

$$X^2 = \sum_{i,j} (a_{ij} - Np_{ij})^2/Np_{ij} \qquad [6]$$

is distributed as χ^2 with $rc - 1$ degrees of freedom.[3] This test was introduced by Pearson (1900). For a proof see Crámer (1946, ch. 30). The quantities Np_{ij} are known in the behavioral science literature as "theoretical frequencies." In this form they are not dependent on the observations except in the sense that $\sum_{ij} Np_{ij} = N$. This one restriction leads to the loss of 1 degree of freedom; hence $df = rc - 1$.

If we consider only the row and column marginals as univariate distributions, then by applying Equation 6 we find

$$X_r^2 = \sum_{i=1}^{r} (a_{i.} - Np_{i.})^2/Np_{i.} \text{ for rows,} \qquad [7]$$

and

$$X_c^2 = \sum_{j=1}^{c} (a_{.j} - Np_{.j})^2/Np_{.j} \text{ for columns.} \qquad [8]$$

As above, X_r^2 and X_c^2 are distributed as χ^2 with $(r - 1)$ and $(c - 1)$ degrees of freedom, respectively.

Now, if as a null hypothesis we assume that there is no association (i.e., independence) between the probability that an observation should occur in a given column and the probability that an observation should occur in a given row, then $p_{ij} = p_{i.}p_{.j}$ for each cell. Hence, we may rewrite Equation 6 under the assumption of independence as

TABLE 1

General Contingency Table

a_{11}	a_{12}	$\cdots$	a_{1j}	$\cdots$	a_{1c}	$a_{1.}$
a_{21}	a_{22}	$\cdots$	a_{2j}	$\cdots$	a_{2c}	$a_{2.}$
$\vdots$			$\vdots$		$\vdots$	$\vdots$
a_{i1}	a_{i2}	$\cdots$	a_{ij}	$\cdots$	a_{ic}	$a_{i.}$
$\vdots$			$\vdots$		$\vdots$	$\vdots$
a_{r1}	a_{r2}	$\cdots$	a_{rj}	$\cdots$	a_{rc}	$a_{r.}$
$a_{.1}$	$a_{.2}$	$\cdots$	$a_{.j}$	$\cdots$	$a_{.c}$	N

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} (a_{ij} - Np_{i.}p_{.j})^2/Np_{i.}p_{.j}, \qquad [9]$$

which is distributed as χ^2 with $rc - 1$ degrees of freedom. Further, under the hypothesis of independence and given the observed marginals only, the maximum likelihood estimate of the expected observed frequency in cell ij is

$$e_{ij} = a_{i.}a_{.j}/N. \qquad [10]$$

Then we may write

$$\begin{aligned} X_k^2 &= \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(a_{ij} - a_{i.}a_{.j}/N)^2}{a_{i.}a_{.j}/N} \\ &\qquad\qquad \text{(for interaction)} \\ &= \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(a_{ij} - e_{ij})^2}{e_{ij}} \qquad [11] \\ &= \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{a_{ij}^2}{e_{ij}} - N. \end{aligned}$$

with $(r - 1)(c - 1)$ degrees of freedom. This component measures the interaction (or association) between the observations. It can be shown (Lancaster, 1949) that the X^2's given by Equations 7, 8, and 11 sum to yield the X^2 given by Equation 9. That is,

$$X^2 = X_r^2 + X_c^2 + X_k^2.$$

Similarly, the degrees of freedom sum to $rc - 1$. That is,

$$rc - 1 = (r - 1) + (c - 1) + (r - 1)(c - 1).$$

Using X_r^2, X_c^2, and X_k^2, it can be seen that the chi-square analysis of contingency tables is an approximation of a trivial case of analysis of

---

[3] The distinction between the quantity X^2 and its distribution is not clearly made in most behavioral science texts. We will observe the more correct procedure of denoting the quantity X^2 and its χ^2 distribution separately.

TABLE 2

General Chi-Square Table

Source	*df*	X^2
Rows	$r-1$	$\sum_{i=1}^{r}(a_{i.}-Np_{i.})^2/Np_{i.}$
Columns	$c-1$	$\sum_{j=1}^{c}(a_{.j}-Np_{.j})^2/Np_{.j}$
Interaction	$(r-1)(c-1)$	$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(a_{ij}-a_{i.}a_{.j}/N)^2}{a_{i.}a_{.j}/N}$
Total	$rc-1$	$\sum_{i=1}^{r}\sum_{j=1}^{c}(a_{ij}-Np_{i.}p_{.j})^2/Np_{i.}p_{.j}$

variance. These results are summarized in Table 2.

Thus far it has been assumed that the marginals of the observed distribution may differ from the theoretical proportions, that is, that $a_{i.}$ and $a_{.j}$ might not or do not have to be equal to $Np_{i.}$ and $Np_{.j}$, respectively. However, it is often the case that the $p_{i.}$ and $p_{.j}$ are not known. When this is the case one uses $a_{i.}$ and $a_{.j}$ as estimates of $Np_{i.}$ and $Np_{.j}$, respectively. If such an assumption is made, then X_r^2 and X_c^2 will be zero. That is, if the row and column marginals are not assumed to be estimates of the population parameters but rather the population parameters themselves, then the row and column sums of squares will be identically zero and all that remains is the row and column interaction sum of squares which is X^2 as usually calculated for a contingency table. We now drop the subscript k from subsequent formulae since all X^2's will be interaction terms.

As already mentioned, the chi-square test assesses the discrepancy between a set of observed frequencies and a set of theoretical frequencies. The larger the discrepancy, the larger the value of chi-square. Intuitively, the larger the discrepancy, the less likely it is that the theoretical distribution could give rise to the obtained distribution. If there are n degrees of freedom, then the χ^2 distribution has the expected value, $E[\chi^2]=n$, and variance $Var[\chi^2]=2n$. Hence, given a value of X^2, one may attach a probability to this obtained value. This probability is the probability that the theoretical proportions could give rise by chance to discrepancies as large or larger than the obtained discrepancy. This probability is a function of both the magnitude of X^2 and the degrees of freedom associated with it.

With the foregoing introduction we turn to specific applications of the chi-square test. Several cases will be considered.

Contingency Tables with 1 Degree of Freedom

Perhaps the most common contingency table is the 2×2 table. As most behavioral science statistics texts have discussions of these tables, detailed discussion will be omitted here. The interested reader is referred to McNemar (1962) and Maxwell (1961) for excellent discussions. Briefly, we consider the variables Y and Z to be dichotomous, or to be categorical, Y may take on values y_1 or y_2, and Z may take on values z_1 or z_2. Then Equation 11 may be simplified in this 2×2 case to

$$X^2=\frac{N(a_{11}a_{22}-a_{12}a_{21})^2}{(a_{11}+a_{12})(a_{11}+a_{21})(a_{21}+a_{22})(a_{12}+a_{22})}, \qquad [12]$$

or if we label the table in the following manner,

		Z	
		z_1	z_2
Y	y_1	A	B
	y_2	C	D

then Equation 12 may be rewritten as

$$X^2=\frac{N(AD-BC)^2}{(A+B)(A+C)(C+D)(B+D)}. \qquad [12a]$$

As was noted earlier, the derivation of th proof that X^2 is distributed as χ^2 is based upo the assumption that the e_{ij}'s, the expecte frequencies in all cells, tend to infinity. Con sequently, if the e_{ij}'s are small, Equation 12 gives rise to inaccurate estimates of the prob bility of the obtained discrepancies. Yate

(1934) has given a "correction for continuity" which should be applied when the e_{ij}'s are small. This correction should be used for *all* 2×2 tables. Using this correction, Equation 12a may be rewritten

$$X^2 = \frac{N(|AD-BC|-N/2)^2}{(A+B)(A+C)(C+D)(B+D)}. \quad [12b]$$

Even with this correction, the values of X^2 do not give the exact probabilities associated with discrepancies. Hence, an even better solution to the small expected value problem is met by calculating the exact probability of the obtained discrepancy. This can be done by use of Fisher's Exact Test which gives the exact probability of occurrence of a set of values, A, B, C, and D, *given fixed marginal frequencies*. Techniques for this procedure are given in most behavioral science statistics texts, for example, McNemar (1962) and Maxwell (1961).

CONTINGENCY TABLES WITH MORE THAN 1 DEGREE OF FREEDOM

Thus far the chi-square test has been discussed with respect to testing hypotheses about independence of proportions. In the 2×2 table inference about independence of proportions was straightforward. However, when the table has more than two rows or two columns the overall chi-square test can only provide information that *somewhere* in the table the observed proportions do not match the theoretical proportions. For most researchers this is not enough. The question arises as to *just where* in a contingency table the important discrepancies occur. If we return to the analysis of variance analogy, we might expect that it is possible to divide X^2 into additive independent components analogous to the orthogonal comparisons or contrasts in the analysis of variance. This can be done (cf. Irwin, 1949; Lancaster, 1949, 1950.) The results provide a very useful method of analyzing contingency tables with more than 1 degree of freedom. Kimball (1954) has derived simple computational formulae similar to those given in the remainder of this paper.

To subdivide X^2 into independent components it is necessary to subdivide the original contingency table in a special manner giving rise to such independent components. If procedures such as those outlined below are not used, then independence of the components is not assured and interpretation is difficult, if not impossible. In the discussion section, further reasons for partitioning will be given. While the partitioning procedures are fixed and standard, it is to be remembered that the ordering of rows and columns in a contingency table is arbitrary. Thus it is necessary for the researcher to order a table so that the resulting partitions are meaningful.

Before giving general partitioning procedures, a few examples should be helpful. Consider a 2×3 table. Such a table can be partitioned into two fourfold tables in the following manner, the numbers under each partition being a number associated with that partition.

$$\underset{(1)}{\begin{array}{c|c} a_{11} & a_{12} \\ \hline a_{21} & a_{22} \end{array}} \qquad \underset{(2)}{\begin{array}{c|c} a_{11}+a_{12} & a_{13} \\ \hline a_{21}+a_{22} & a_{23} \end{array}}.$$

Note that all frequencies occurring in a cell of a partition are to be summed. For instance, in Partition 2 above, the frequency in the upper left-hand cell is $a_{11} + a_{12}$, the frequency in the lower left is $a_{21} + a_{22}$, etc. For a 3×3 table the partitions are

$$\underset{(1)}{\begin{array}{c|c} a_{11} & a_{12} \\ \hline a_{21} & a_{22} \end{array}} \qquad \underset{(2)}{\begin{array}{c|c} a_{11}+a_{12} & a_{13} \\ \hline a_{21}+a_{22} & a_{23} \end{array}}$$

$$\underset{(3)}{\begin{array}{c|c} a_{11} & a_{12} \\ + & + \\ a_{21} & a_{22} \\ \hline a_{31} & a_{32} \end{array}} \qquad \underset{(4)}{\begin{array}{c|c} a_{11}+a_{12} & a_{13} \\ + & + \\ a_{21}+a_{22} & a_{23} \\ \hline a_{31}+a_{32} & a_{33} \end{array}}.$$

The general contingency table (Table 1) can be partitioned into $(r-1)(c-1)$ fourfold tables with 1 degree of freedom each. The procedure for generating the partitions is quite simple. For such a table, a partition is formed by taking the last $r-1$ row and $c-1$ column frequencies as the lower right member of the $(r-1)(c-1)$ independent fourfold tables. The remaining three cell frequencies for each fourfold table are obtained as follows:

The lower left cell contains the row sum to the left of the lower right frequency, the upper right cell contains the column sum above the lower right frequency, the upper left cell frequency is the sum of all frequencies above and to the left of the lower right frequency. That is, the partition associated with cell a_{ij} can be represented by the following fourfold table:

$$\begin{array}{ll|l} a_{11}+a_{12}\cdot & \cdot\cdot a_{1,j-1} & a_{1j} \\ a_{21}+a_{22}\cdot & \cdot\cdot & a_{2j} \\ \vdots & \vdots & \vdots \\ a_{i-1,1} & \cdot\cdot\cdot a_{i-1,j-1} & a_{i-1,j} \\ \hline a_{i1} & \cdot\cdot\cdot a_{i,j-1} & a_{ij} \end{array} .$$

This procedure ensures that independent subdivisions of the original contingency table have been made and facilitates computation of the component X^2s.

$2 \times c$ *Tables*

Formulae. For the 2×3 table above, the X^2s associated with the two partitions are

$$X^2{}_1 = \frac{N^2(a_{11}a_{22}-a_{12}a_{21})^2}{a_{1.}a_{2.}a_{.1}a_{.2}(a_{.1}+a_{.2})}, \quad \text{and} \qquad [13]$$

$$X^2{}_2 = \frac{N[a_{23}(a_{11}+a_{12})-a_{13}(a_{21}+a_{22})]^2}{a_{1.}a_{2.}a_{.3}(a_{.1}+a_{.2})}, \qquad [14]$$

the subscript t on the $X^2{}_t$ referring to the tth partition.

In general, for a $2 \times c$ table, the partitioned X^2s can be computed from

$$X^2{}_t = \frac{N^2[a_{2,t+1}\sum_{j=1}^{t} a_{1j} - a_{1,t+1}\sum_{j=1}^{t} a_{2j}]^2}{a_{1.}a_{2.}a_{.,t+1}[\sum_{j=1}^{t} a_{.j}][\sum_{j=1}^{t+1} a_{.j}]}, \qquad [15]$$

where $t = 1, 2, \cdots, (c-1)$. Each of these X^2s can be summed to yield the overall X^2, that is,

$$X^2 = \sum_{t=1}^{c-1} X^2{}_t.$$

Example of a $2 \times c$ *Table.* Suppose we have data on some variable Y for four ethnic groups. Let Y be high or low. Then we can make the following table:

TABLE 3

Example of $2 \times c$ Table

Variables	Rural Indian	Local Indian	Spanish	Anglo	Totals
High Y	21	8	24	13	66
Low Y	14	3	12	30	59
Totals	35	11	36	43	125

If we calculate X^2 for the entire table using Equation 11, we obtain $X^2 = 14.03$, and with 3 degrees of freedom, reference to a χ^2 distribution table indicates a significance probability of $p < .005$. This large X^2 and small significance probability indicate that there is interaction between ethnic status and the variable Y. That is, there is a large discrepancy between the observed frequencies and the theoretical frequencies as determined by the marginal frequencies. However, it may be that the discrepancy is large for some cells and small for others. In order to discover where in the table the major discrepancies arise, we resort to the partitioning procedures described for $2 \times c$ tables. As there are 3 degrees of freedom we may obtain three partitions. These three partitions allow comparisons to be made as follows: for the Rural Indian-Local Indian (RI × LI) comparison,

$$X^2{}_1 = \frac{125^2(3\cdot 21 - 8\cdot 14)^2}{66\cdot 59\cdot 11\cdot 35\cdot 46} = 0.54,$$

and for the Indian-Spanish [(RI − LI) × S] comparison,

$$X^2{}_2 = \frac{125^2(12\cdot 29 - 24\cdot 17)^2}{66\cdot 59\cdot 36\cdot 46\cdot 82} = 0.10,$$

and for the nonAnglo-Anglo [(RI−LI−S)×A] comparison,

$$X^2{}_3 = \frac{125^2(30\cdot 53 - 13\cdot 29)^2}{66\cdot 59\cdot 43\cdot 82\cdot 125} = 13.39.$$

These partitioned X^2s may now be presented in an "X^2 Table."

This analysis indicates that there is no variation within the nonAnglo groups while there is high (and significant) variation between nonAnglos and Anglos.

TABLE 4

X^2 TABLE FOR $2 \times c$ EXAMPLE

Source	*df*	X^2	*p*
(1) RI×LI	1	.54	*ns*
(2) (RI−LI)×S	1	.10	*ns*
(3) (RI−LI−S)×A	1	13.39	$p<.001$
Total	3	14.03	$p<.005$

$r \times c$ *Tables*

Formulae. For the 3×3 table given above the X^2s associated with each partition are

$$X^2_1 = \frac{N[a_{2.}(a_{.2}a_{11} - a_{.1}a_{12}) - a_{1.}(a_{.2}a_{21} - a_{.1}a_{22})]^2}{a_{1.}a_{2.}a_{.1}a_{.2}(a_{1.}+a_{2.})(a_{.1}+a_{.2})}, \quad [16]$$

$$X^2_2 = \frac{N^2[a_{23}(a_{11}+a_{12}) - a_{13}(a_{21}+a_{22})]^2}{a_{1.}a_{2.}a_{.3}(a_{1.}+a_{2.})(a_{.1}+a_{.2})}, \quad [17]$$

$$X^2_3 = \frac{N^2[a_{32}(a_{11}+a_{21}) - a_{31}(a_{12}+a_{22})]^2}{a_{3.}a_{.1}a_{.2}(a_{1.}+a_{2.})(a_{.1}+a_{.2})}, \quad [18]$$

$$X^2_4 = \frac{N[a_{33}(a_{11}+a_{12}+a_{21}+a_{22}) - (a_{13}+a_{23})(a_{31}+a_{32})]^2}{a_{3.}a_{.3}(a_{1.}+a_{2.})(a_{.1}+a_{.2})}. \quad [19]$$

In general, for an $r \times c$ table the partitioned X^2 associated with a_{ij} is given by

$$X^2_t = \frac{N\left[a_{.j}\left(a_{i.}\sum_{h=1}^{i-1}\sum_{k=1}^{j-1} a_{hk} - \sum_{h=1}^{i-1} a_{h.} \sum_{k=1}^{j-1} a_{ik}\right) - \sum_{k=1}^{j-1} a_{.k}\left(a_{i.}\sum_{h=1}^{i-1} a_{hj} - a_{ij}\sum_{h=1}^{i-1} a_{h.}\right)\right]^2}{a_{.j}a_{i.}\left(\sum_{k=1}^{j-1} a_{.k}\right)\left(\sum_{k=1}^{j} a_{.k}\right)\left(\sum_{h=1}^{i-1} a_{h.}\right)\left(\sum_{h=1}^{i} a_{h.}\right)}$$

where $t = (c-1)(i-2) + j - 1;\ i = 2, \cdots, r$ and $j = 2, \cdots, c$. It should be noted that the general formulae are not as formidable as might first appear. After computing only a few partitioned X^2s, one can readily apply the method without recourse to the formulae.

Example. Returning to the study of our ethnic groups, suppose that we have three ethnic groups, and suppose further that we have a measure on a trichotomous variable which measures an attitude *Y*. Responses to the attitude measure can be negative (−), neutral (0), or positive (+). The data are presented in Table 5.

TABLE 5

EXAMPLE OF $r \times c$ TABLE

Group	Attitude score			
	−	0	+	Totals
Indian	25	17	20	62
Spanish	27	7	25	59
Anglo	13	33	45	91
Totals	65	57	90	212

If we calculate X^2 for the entire table using Equation 11, we obtain $X^2 = 24.75$, and with 4 degrees of freedom, $p < .0001$. This large and significant X^2 indicates that there is interaction between ethnic status and the measured attitude *Y*. To study this interaction in greater detail we partition X^2.

The partitioned X^2s are as follows: (1) For the Indian versus Spanish comparison on negative versus neutral attitude,

$$X^2_1 = \frac{212[59(57\cdot 25 - 65\cdot 17) - 62(57\cdot 27 - 65\cdot 7)]^2}{62\cdot 59\cdot 65\cdot 57\cdot 121\cdot 122} = 2.47,$$

(2) for the Indian versus Spanish comparison on negative and neutral attitude versus positive attitude,

$$X^2_2 = \frac{212^2(25\cdot 42 - 20\cdot 34)^2}{62\cdot 59\cdot 90\cdot 121\cdot 122} = 1.27,$$

(3) for the nonAnglo versus Anglo comparison on negative versus neutral attitude,

$$X^2_3 = \frac{212^2(33\cdot 52 - 13\cdot 24)^2}{91\cdot 65\cdot 57\cdot 121\cdot 122} = 17.81,$$

(4) finally, for the nonAnglo versus Anglo comparison in negative and neutral attitude versus positive attitude,

$$X^2_4 = \frac{212(45\cdot 76 - 45\cdot 46)^2}{91\cdot 90\cdot 121\cdot 122} = 3.20.$$

These statistics may be summarized in a X^2 Table.

These data suggest that while there are overall ethnic group differences on the attitude

TABLE 6

X^2 TABLE FOR THE $r \times c$ EXAMPLE

Source	df	X^2	p
(1) I×S on (– vs. 0)	1	2.47	*ns*
(2) I×S on (–, 0 vs. +)	1	1.27	*ns*
(3) (I–S)×A on (– vs. 0)	1	17.81	$p<.0001$
(4) (I–S)×A on (–, 0 vs. +)	1	3.20	$p<.10$
Total	4	24.75	$p<.0001$

Y, there are no differences between the Indian and Spanish groups. Consequently, the high X^2 is primarily due to differences in the Anglo and nonAnglo cells. Further, most of these differences are due to negative and neutral attitudes with only slight differences among Anglos and nonAnglos due to the comparison of negative-neutral attitudes and positive attitude. Note that exchange of rows and columns may be necessary to arrange the table to test other specific comparisons with these procedures.

A Special Partition of 2 × c Tables

Procedure. There is yet another useful method of partitioning a $2 \times c$ table which is of interest to the behavioral scientist. This method is due to Brandt and Snedecor and is somewhat more familiar. Suppose there are c groups (or c values of some variable) for which some dichotomous variable is measured (e.g., the presence or absence of some trait) for the several groups. The data could be represented as in Table 7.

Then X^2 for this table can be given by

$$X^2 = \frac{N}{a_{1.}a_{2.}}\left[N \sum_{j=1}^{c} \frac{a_{1j}^{\,2}}{a_{.j}} - a_{1.}^{\,2}\right] \qquad [21]$$

TABLE 7

GENERAL REPRESENTATION OF SPECIAL PARTITION OF $2 \times c$ TABLE

Trait	Group				Totals
	1	2	$\cdots$	c	
Present	a_{11}	a_{12}	$\cdots$	a_{1c}	$a_{1.}$
Absent	a_{21}	a_{22}	$\cdots$	a_{2c}	$a_{2.}$
Totals	$a_{.1}$	$a_{.2}$	$\cdots$	$a_{.c}$	N

with $c-1$ degrees of freedom. Note that Equation 21 is equivalent to Equation 11.

Suppose we believe that for the c groups there is some a priori basis for hypothesizing that the first c_1 groups are different from the subsequent c_2 groups on the basis of a third variable, say sex, age, education, etc., where the groups may be ordered in any arbitrary manner. It would be desirable to test for differences between these two sets of groups. One way to do this would be to collapse the two sets of groups into two groups and test for differences in the resulting fourfold table. However, by doing this valuable information would be lost, especially if the a priori assumption about the groups should be incorrect. By partitioning we can assess the homogeneity within and between the two sets of groups in the following manner. The X^2 given by Equation 21 can be partitioned into three additive components:

X^2 due to variation within the set of c_1 groups,
X^2 due to variation within the set of c_2 groups,
X^2 due to the variation between the two groups.

The degrees of freedom associated with these X^2s are $c_1 - 1$, $c_2 - 1$, and 1, respectively. These three X^2s can be added to obtain the X^2 calculated for the entire table.

The procedure is to alter Table 7 into two tables:

TABLE 8

GENERAL REPRESENTATION OF SPECIAL PARTITION OF $2 \times c$ TABLE FOR TWO SETS OF GROUPS

Trait	C₁ groups				Totals
	1	2	$\cdots$	c_1	
Present	a_{11}	a_{22}	$\cdots$	a_{1c_1}	$a_{1.,c_1}$
Absent	a_{21}	a_{22}	$\cdots$	a_{2c_1}	$a_{2.,c_1}$
Totals	$a_{.1}$	$a_{.2}$	$\cdots$	$a_{.c_1}$	N_{c_1}

Trait	C₂ groups			Totals
	$c_1 + 1$	$\cdots$	c	
Present	a_{1,c_1+1}	$\cdots$	a_{1c}	$a_{1.,c_2}$
Absent	a_{2,c_1+1}	$\cdots$	a_{2c}	$a_{2.,c_2}$
Totals	$a_{.,c_1+1}$	$\cdots$	$a_{.c}$	N_{c_2}

Note that $N = N_{c_1} + N_{c_2}$,

$$a_{1.,c_1} + a_{1.,c_2} = a_{1.}, \text{ etc.}$$

Then the computational formulae for the partitions become:

$$X^2_{c_1} = \frac{N_{c_1}}{a_{1.,c_1}a_{2.,c_1}}\left[N_{c_1}\sum_{j=1}^{c_1}\frac{a_{1j}^2}{a_{.j}} - a_{1.,c_1}^2\right], \qquad [22]$$

$$X^2_{c_2} = \frac{N_{c_2}}{a_{1.,c_2}a_{2.,c_2}}\left[N_{c_2}\sum_{j=c1+1}^{c}\frac{a_{1j}^2}{a_{.j}} - a_{1.,c_2}^2\right], \quad [23]$$

$$X^2_{c_1\times c_2} = \frac{N^2}{a_{1.}a_{2.}}\left[\frac{a_{1.,c_1}^2}{N_{c_1}} + \frac{a_{1.,c_2}^2}{N_{c_2}} - \frac{a_{1.}^2}{N}\right]. \quad [24]$$

Example. Suppose we have the following data on some trait Y for three ethnic groups. However, we wish to know if there are any sex differences when we also expect ethnic group differences. We may represent the data in Table 9.

TABLE 9

EXAMPLE OF SPECIAL PARTITION OF 2 × *c* TABLE

Trait Y	Groups						Totals
	A♂	S♂	I♂	A♀	S♀	I♀	
Present	6	11	13	6	15	16	67
Absent	17	4	11	14	6	6	58
Totals	23	15	24	20	21	22	125

Then, using Formulae 21 through 24, we have

$$X^2 = \frac{125}{65\cdot 58}[125\cdot 40.824 - 67^2] = 19.79$$

for the overall X^2, and

$$X^2_m = \frac{62}{30\cdot 32}[62\cdot 16.674 - 30^2] = 8.60;$$

$$X^2_f = \frac{63}{37\cdot 26}[63\cdot 24.151 - 37^2] = 9.95;$$

$$X^2_{m\times f} = \frac{125^2}{67\cdot 58}\left[\frac{30^2}{62} + \frac{37^2}{63} - \frac{67^2}{125}\right] = 1.31,$$

where *m* indicates the male group and *f* indicates the female group. These results may be summarized in an X^2 Table.

TABLE 10

X^2 TABLE FOR EXAMPLE OF SPECIAL PARTITION OF 2 × *c* TABLE

Source	*df*	X^2	*p*
Within males	2	8.60	$p<.02$
Within females	2	9.95	$p<.01$
Males × Females	1	1.31	*ns*
Total	5	19.79	$p<.005$

These results allow us to infer that there are ethnic group differences on trait Y for both males and females, but that there are no overall sex differences on trait Y—hence it is ethnicity, not sex, which leads to overall significance in the above 2 × 6 contingency table.

Small Expected Values in the r × c Table

As was mentioned for the 2 × 2 table certain corrections for the X^2 value in such a table are warranted on the basis of small expected frequencies. When a table is not 2 × 2 it would seem that some correction would be called for when small expected frequencies are involved. Usually the procedure has been to collapse (pool) categories with small expected frequencies. Cochran (1954) has shown that such a procedure seems much too conservative and advises against the traditional pooling criteria, as does Maxwell (1961). For contingency tables with small N or few rows and few columns the researcher can resort to exact probability test procedures given by Freeman and Halton (1951) and Bennett and Nakamura (1963). For further discussion see these four references.

With respect to partitioning procedures, pooling of small expected frequencies and corrections for continuity should not be done.

DISCUSSION

At this point, there may still be questions as to why the researcher should resort to the rather complicated partitioning procedures when it would be easier to set up several 2 × 2 tables and calculate the X^2 for each one using Equation 12b. There are two objections to doing this.

First, recall that in setting up the contingency tables it was assumed that the marginal observed frequencies were the population values. This being so, the marginals of any arbitrary 2 × 2 table extracted from a larger table may not reflect the true marginals of the larger table, since any 2 × 2 table on which X^2 is computed from Equation 12a or 12b assumes that the population is bifurcated in the proportions indicated by the marginal values of that table. That is, any selected 2 × 2 tables would reflect the nature of as many populations as tables. Recall that the categories must be *both* mutually exclusive and exhaustive of the population on which inference is desired. The method of partitioning the degrees of freedom allows generalization back to the *entire* population and not just to the population contained in the partition. Thus in the example given first, inference about the ethnic groups could be related back to a single population containing all of the ethnic groups. Hence we were able to reject an hypothesis that the two Indian groups were dissimilar in a population containing *all* of the ethnic groups.

Second, as discussed earlier, the X^2s for such arbitrary 2 × 2 tables would be correlated, and appropriate inferences on a set of such X^2s would prove to be difficult. However, partitioned X^2s are not correlated and can be interpreted *independently*, thus allowing meaningful inference.

It is important to remember that the number of partitions in a contingency table is limited by the degrees of freedom and that once a table has been laid out the partitions are dictated by the procedures outlined above. As the cells may be ordered in many different ways, the researcher is able to lay out his table a priori in any way which is desirable. However, any layout for partitions should be determined before calculating any of the partitioned X^2s. For instance, in the first example it was hypothesized that there was a degree of similarity among nonAnglos and that the non-Anglo groups should differ from the Anglo group. Consequently, the table was laid out so that the Rural Indians and Local Indians could be compared and similarity among Indian groups assessed. It was then possible to assess differences between Spanish and Indians. Lastly, the differences between non-Anglos and Anglos were evaluated. In this manner the relevant hypotheses were tested.

REFERENCES

Bennett, B. M., & Nakamura, E. Tables for testing significance in a 2 × 3 table. *Technometrics*, 1963, **5**, 501–511.

Cochran, W. G. Some methods for strengthening the common chi-square tests. *Biometrics*, 1954, **10**, 417–451.

Cramér, H. *Mathematical methods of statistics*. Princeton: Princeton Univer. Press, 1946.

Freeman, G. H., & Halton, J. H. Note on an exact treatment of contingency, goodness-of-fit, and other problems of significance. *Biometrika*, 1951, **38**, 141–149.

Irwin, J. O. A note on the subdivision of chi-square into components. *Biometrika*, 1949, **36**, 130–134.

Kimball, A. W. Short-cut formulas for the exact partition of chi-square in contingency tables. *Biometrics*, 1954, **10**, 452–458.

Lancaster, H. O. The derivation and partition of chi-square in certain discrete distributions. *Biometrika*, 1949, **36**, 117–129.

Lancaster, H. O. The exact partition of chi-square and its application to the problem of the pooling of small expectations. *Biometrika*, 1950, **37**, 267–270.

Maxwell, A. E. *Analysing qualitative data*. London: Methuen, 1961.

McNemar, Q. *Psychological statistics*. (3rd ed.) New York: Wiley, 1962.

Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 1900, **5** (50) 157–175.

Wilson, K. V. A distribution-free test of analyses of variance hypotheses. *Psychological Bulletin*, 1956, **53**, 96–101.

Yates, F. Contingency tables involving small numbers and the chi-square test. *Journal of the Royal Statistical Society*, 1934, **1** (Suppl.), 217–235.

(Received August 24, 1964)

Psychological Bulletin
1965, Vol. 64, No. 5, 339–350

DESIGN OF EXPERIMENTS TO MAXIMIZE POWER RELATIVE TO COST[1]

JOHN E. OVERALL AND SUDHIR N. DALAL

University of Texas Medical Branch

The relationship of power of F tests to expected mean squares, E(MS), in the analysis of variance is discussed. While components of variance in the E(MS) are largely a function of nature, the coefficients associated with them are matters of experimental design. Frequently a different cost is associated with each type of experimental unit represented by the different coefficients. It is possible to maximize power relative to cost by optimal allocation of available resources among various types of experimental units—for example, numbers of Ss, duplicate measures, replicates, etc. A simple index of relative power involving the ratio of the estimated F ratio to F_α is proposed as useful in choosing the allocation of resources most likely to yield significant results.

A major contribution by the statistically trained psychologist should be in the design of experiments. In particular, it is important to allocate resources in such a way that there is a maximum probability of detecting true treatment differences when they exist. Consideration of expected mean squares, *E*(*MS*), and estimates of components of variance makes such a logical choice of experimental design a practicality in many situations.

Psychologists are familiar with the use of *E*(*MS*) for selection of proper error terms in the analysis of variance (Gaito, 1961). Methods of deriving *E*(*MS*) have been presented in various texts (Federer, 1955; Lindquist, 1953; Winer, 1962), and, in fact, complete tables of *E*(*MS*) values for most experimental designs are provided in these and other sources. In general, a proper *F* ratio is one which contains in the denominator all components of variance contained in the numerator except the one associated with the parameter being tested.

In most experimental situations, the *E* (experimenter) has various alternatives available to him in efforts to maximize power in testing hypotheses. He can increase sample size, use more reliable measurement techniques, secure multiple measurement on each *S* (subject), stratify samples into larger or smaller number of levels according to values on one or more control variables, or do one of a number of other things which will tend to reduce error relative to treatment components. Unless he has unlimited resources, he must weigh the probable value of these alternative methods against their costs. If his resources are quite limited, he may wish to maximize the probability of obtaining a significant *F* ratio relative to the cost of the experiment.

The purpose of this article is to discuss the usefulness of *E*(*MS*) and components of variance estimates in designing experiments to have maximum power relative to cost. The probability that an *F* ratio will result in rejection of the null hypothesis (when it is really false) depends upon the alpha level, degrees of freedom, and upon the components of variance and their associated coefficients in the *E*(*MS*). While the components of variance representing various error factors may be largely a function of nature, the magnitudes of coefficients associated with them are experimental parameters under the control of *E*. These coefficients represent the number of *S*s, number of levels of a control variable, and so on. If estimates of components of variance are available, *E* is free to design the experiment so as to maximize the probability of a significant *F* ratio within the limitation of resources available to him.

E(*MS*) *and F Ratios.* The *F* ratio used in testing a statistical hypothesis contains in its denominator all components of variance included in the numerator except the one related to the *H* being tested.[2] If the component of variance due to treatment effect being tested

---

[1] This research was accomplished using a computer facility made available by Special Research Resources Grant 1 PO 7 FR 00024–01, National Institutes of Health.

[2] This use of the term "component of variance" in connection with fixed treatment effects is slightly paradoxical; however, the meaning is clear and the variance of the fixed treatment is $\sigma_t^2 = \Sigma B_i^2/k$.

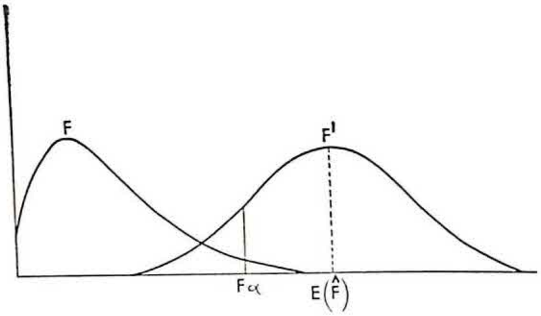


FIG. 1. Heuristic diagram of central and noncentral F distributions.

is actually equal to zero, and the numerator and denominator are independently distributed as chi-squares, then the sampling distribution of the ratio will be a "central F distribution." It is frequently said that the central F distribution is the sampling distribution of the ratio of two independent estimates of the same population variance, and several writers have shown that the central F distribution provides an adequate approximation, even if numerator and denominator do not have chi-square distributions, so long as both are independent estimates of the same population variance (Satterthwaite, 1941). On the other hand, if the numerator of the ratio contains an additional nonzero treatment term, the numerator and denominator will not be estimates of the same population variance and the statistic will have a "noncentral F distribution."

$$F = \frac{\sigma_o^2 + m\sigma_t^2}{\sigma_o^2}, \qquad [1]$$

where σ_o^2 is the total variance common to both the numerator and denominator, and $m\sigma_t^2$ is the additional term present in the numerator but not the denominator.

In the general case, the denominator and numerator may contain several common terms, that is, the error variance may be composed of several different components. Let σ_o^2 be the general representation for all components of variance appearing in both the numerator and denominator of the proper F ratio. In a similar manner, let $m\sigma_t^2$ represent the addition treatment component where m stands for magnitude and is a general notation for all coefficients associated with that component of variance. As one specific example, consider the F ratio for testing the significance of the main effect for A in a two-way mixed-model analysis of variance:

$$F = \frac{\sigma_\epsilon^2 + n\sigma_{\mathsf{AB}}^2 + bn\sigma_{\mathsf{A}}^2}{\sigma_\epsilon^2 + n\sigma_{\mathsf{AB}}^2}. \qquad [2]$$

In this example, the total error variance is composed of two components:

$$\sigma_o^2 = \sigma_\epsilon^2 + n\sigma_{\mathsf{AB}}^2.$$

The treatment component σ_{A}^2 has coefficients bn so that $bn = m$ of Equation 1. In a similar manner, the variance in any F ratio can be conceived in terms of error and treatment components as in Equation 1.

In Figure 1 is presented a heuristic diagram showing theoretical sampling distribution of F under two conditions. The distribution marked F is the *central* F distribution which includes the range of values one would expect if indeed the statistic computed was the ratio of two independent estimates of the same population variance—if $m\sigma_t^2 = 0$. A value F_α is chosen as being the largest value likely to occur if indeed the statistic computed is the ratio of two independent estimates of the same variance. If an F ratio having a value larger than F_α is obtained, it is concluded that the ratio probably did not belong to the central F distribution. The distribution marked F' is the theoretical sampling distribution of an F ratio in which the numerator contains a nonzero treatment component $m\sigma_t^2$ not present in the denominator. The distribution F' is the sampling distribution which the F statistic would actually have if the null hypothesis were false and if the magnitude of true treatment differences were $m\sigma_t^2$.

In Figure 1, the value F_α is the critical value of F beyond which only alpha percent (say, .05 or .01) of the total range of values would be expected to fall if indeed the null hypothesis were true. If the null hypothesis is really false, then the F ratio will contain a nonzero treatment component in the numerator and will have a noncentral F' distribution. In the F' distribution, the probability of an observed F ratio exceeding F_α is equal to the area under F' lying to the right of F_α. The probability of making a Type II error (failing to reject H_o when false) is equal to the area under F' to the left of F_α. In order to calculate the power of a test, we need to know the distribution F' and the critical value F_α. The power of the test is then the integral of F' from F_α to zero.

The distribution F' depends on degrees of freedom (df) and on the noncentrality parameter λ (Fisher, 1928; Patnaik, 1949). Various writers have presented power function curves (Feldt & Mahmoud, 1958; Pearson & Hartley, 1951), but in general these are not satisfactory for the applications considered here. Power function values calculated for an extended set of parameter values are presented in Table 1. The values in the body of this table are the power of F tests for α .05, df N_1 and N_2, and noncentrality parameter λ.

The noncentrality parameter λ can be calculated directly from estimates of components of variance in the numerator and denominator of the F ratio for the test under consideration. From the general form of Equation 1, the noncentrality parameter is as follows:

$$\lambda = km\sigma_t^2/\sigma_o^2, \qquad [3]$$

where k is the number of treatment means entering into the null hypothesis, σ_t^2 is the variance of true treatment means, and m is the total coefficient associated with σ_t^2 in the $E(MS)$, and where σ_o^2 is the total appropriate error variance. In the general case, the denominator of the F ratio will contain several components of variance—that is, the total error variance, σ_o^2, will be composed of several components. In a similar manner, the coefficient m may be a product of several coefficients associated with the treatment component in the $E(MS)$ of the numerator. For example, in the two-way analysis of Equation 2, the noncentrality parameter λ would be as follows:

$$\lambda = kbn\sigma_{\mathrm{A}}^2/(\sigma_\epsilon^2 + n\sigma_{\mathrm{AB}}^2). \qquad [4]$$

In a similar manner, *the* λ *value for any* F *test contains in the numerator* k *times the treatment term, including coefficients, from the numerator of the expected* F *ratio, and the denominator of* λ *contains all components of variance with coefficients appearing in the* $\mathrm{E(MS)}_\epsilon$. The coefficient k is the number of means entering into the hypothesis. In the case of a test of interaction, k will be the number of *cell* means involved.

Obtaining Components of Variance Estimates. Whenever previous research results in an analysis of variance summary table, it is possible to compute estimates of all components of variance appearing in the $E(MS)$. The first step is to write the complete set of $E(MS)$ for the particular experimental design. Suppose, for example, that the design is a hierarchal design in which there are m independent normally distributed observations on each S and in which there are n different Ss in each treatment group. The $E(MS)$s for this design, as given in Winer (1962, p. 185) and in other sources, are as follows:

$$E(MS_t) = \sigma_\epsilon^2 + m\sigma_s^2 + nm\sigma_t^2$$
$$E(MS_s) = \sigma_\epsilon^2 + m\sigma_s^2$$
$$E(MS_\epsilon) = \sigma_\epsilon^2.$$

For any table of $E(MS)$s, the estimates of individual components of variance can be calculated from the MSs obtained in an analysis of variance. The best unbiased estimate of the error component is the MS for error:

$$MS_\epsilon = \sigma_\epsilon^2.$$

Substituting this value into each equation containing only this value plus one other component, an estimate of the component of variance associated with the next level can be computed. In the hierarchal design of this example, the component of variance associated with variation of observations within Ss can be estimated by substituting $MS_\epsilon = \sigma_\epsilon^2$ into the equation for $E(MS_s)$:

$$MS_s = \sigma_\epsilon^2 + m\sigma_s^2$$
$$MS_s = MS_\epsilon + m\sigma_s^2$$
$$\sigma_s^2 = \frac{MS_s - MS_\epsilon}{m}.$$

Substituting the two estimates into each equation containing these two plus one other yields estimates of components of variance at the next level. In the example considered here, the component of variance due to treatment differences can be estimated as follows:

$$MS_t = \sigma_\epsilon^2 + m\sigma_s^2 + nm\sigma_t^2$$
$$MS_t = MS_\epsilon + m\left(\frac{MS_s - MS_\epsilon}{m}\right) + nm\sigma_t^2$$
$$MS_t = MS_s + nm\sigma_t^2$$
$$\sigma_t^2 = \frac{MS_t - MS_s}{nm}.$$

TABLE 1

Probability of Rejecting Null Hypothesis at .05 Level

N_1	N_2	$\lambda=3$	$\lambda=4$	$\lambda=5$	$\lambda=6$	$\lambda=8$	$\lambda=10$	$\lambda=12$	$\lambda=18$
1	2	.18	.22	.26	.29	.36	.42	.47	.61
1	3	.23	.29	.34	.40	.49	.58	.65	.80
1	4	.26	.33	.39	.46	.57	.67	.74	.89
1	5	.28	.37	.43	.51	.63	.72	.80	.93
1	6	.30	.39	.46	.54	.66	.76	.83	.95
1	8	.32	.42	.49	.58	.71	.80	.87	.97
1	10	.33	.44	.51	.60	.73	.83	.89	.98
1	12	.34	.45	.52	.62	.75	.84	.90	.98
1	18	.36	.47	.54	.65	.78	.87	.92	.99
1	30	.37	.49	.56	.67	.80	.89	.94	.99
1	60	.38	.50	.58	.68	.81	.90	.95	.99
2	2	.12	.14	.16	.18	.22	.26	.30	.39
2	3	.15	.19	.22	.25	.32	.38	.44	.59
2	4	.18	.22	.26	.31	.39	.47	.54	.71
2	5	.19	.24	.29	.35	.44	.53	.61	.79
2	6	.21	.26	.32	.38	.49	.58	.67	.84
2	8	.23	.30	.36	.42	.54	.65	.73	.89
2	10	.24	.31	.38	.45	.58	.68	.77	.92
2	12	.25	.33	.40	.47	.60	.71	.79	.94
2	18	.27	.35	.43	.51	.65	.75	.83	.96
2	30	.28	.37	.45	.54	.68	.79	.86	.97
2	60	.30	.39	.47	.56	.70	.81	.88	.98
3	2	.10	.11	.13	.14	.17	.20	.22	.30
3	3	.12	.14	.16	.19	.23	.28	.32	.45
3	4	.14	.17	.20	.23	.30	.36	.42	.58
3	5	.15	.19	.23	.27	.34	.42	.49	.66
3	6	.16	.21	.25	.29	.38	.46	.54	.73
3	8	.18	.23	.28	.33	.43	.53	.61	.80
3	10	.19	.25	.31	.36	.47	.57	.66	.84
3	12	.20	.26	.32	.38	.49	.60	.69	.87
3	18	.22	.29	.35	.42	.54	.65	.74	.91
3	30	.24	.31	.38	.45	.58	.69	.78	.94
3	60	.25	.33	.40	.47	.61	.72	.81	.95
4	2	.08	.10	.11	.12	.14	.16	.18	.24
4	3	.10	.12	.14	.16	.20	.23	.34	.37
4	4	.12	.14	.16	.19	.24	.29	.41	.48
4	5	.13	.16	.19	.22	.28	.34	.46	.57
4	6	.14	.17	.21	.24	.31	.39	.46	.64
4	8	.15	.20	.24	.28	.36	.45	.53	.72
4	10	.16	.21	.26	.31	.40	.49	.58	.78
4	12	.17	.22	.27	.33	.43	.52	.61	.81
4	18	.19	.25	.30	.36	.48	.58	.67	.86
4	30	.21	.27	.33	.40	.52	.63	.73	.90
4	60	.22	.29	.36	.43	.56	.67	.77	.93
5	2	.08	.09	.10	.11	.12	.14	.16	.21
5	3	.09	.11	.12	.14	.17	.20	.23	.32
5	4	.10	.12	.14	.16	.21	.25	.29	.41
5	5	.11	.14	.16	.19	.24	.29	.35	.50
5	6	.12	.15	.18	.21	.27	.33	.39	.56
5	8	.13	.17	.20	.24	.32	.39	.46	.65
5	10	.14	.19	.22	.27	.35	.43	.51	.71
5	12	.15	.20	.24	.28	.38	.46	.55	.75
5	18	.17	.22	.27	.32	.43	.53	.62	.82
5	30	.19	.24	.30	.36	.47	.58	.68	.87
5	60	.20	.26	.32	.39	.51	.63	.72	.91

Table 1—Continued

N_1	N_2	$\lambda=3$	$\lambda=4$	$\lambda=5$	$\lambda=6$	$\lambda=8$	$\lambda=10$	$\lambda=12$	$\lambda=18$
6	2	.07	.08	.09	.10	.11	.13	.14	.18
6	3	.09	.10	.11	.12	.15	.17	.20	.28
6	4	.09	.11	.13	.15	.18	.22	.25	.36
6	5	.10	.12	.14	.17	.21	.26	.30	.44
6	6	.11	.13	.16	.18	.24	.29	.34	.50
6	8	.12	.15	.18	.21	.28	.34	.41	.59
6	10	.13	.17	.20	.24	.31	.38	.46	.65
6	12	.14	.18	.21	.25	.33	.42	.50	.70
6	18	.15	.20	.24	.29	.39	.48	.57	.78
6	30	.17	.22	.27	.33	.43	.54	.63	.84
6	60	.18	.24	.30	.36	.48	.59	.69	.88
8	2	.07	.07	.08	.08	.10	.11	.12	.15
8	3	.08	.09	.10	.11	.12	.14	.16	.22
8	4	.08	.10	.11	.12	.15	.18	.20	.29
8	5	.09	.11	.12	.14	.17	.21	.24	.35
8	6	.10	.11	.13	.15	.19	.23	.28	.40
8	8	.10	.13	.15	.17	.22	.28	.33	.49
8	10	.11	.14	.17	.19	.25	.31	.38	.55
8	12	.12	.15	.18	.21	.28	.34	.41	.60
8	18	.13	.17	.21	.24	.32	.40	.49	.69
8	30	.14	.19	.23	.28	.37	.46	.55	.77
8	60	.16	.21	.26	.31	.42	.52	.62	.82
10	2	.06	.07	.07	.08	.09	.10	.11	.13
10	3	.07	.08	.09	.09	.11	.13	.14	.19
10	4	.08	.09	.10	.11	.13	.15	.17	.24
10	5	.08	.09	.11	.12	.15	.17	.20	.29
10	6	.09	.10	.12	.13	.16	.20	.23	.34
10	8	.09	.11	.13	.15	.19	.24	.28	.42
10	10	.10	.12	.14	.17	.22	.27	.32	.48
10	12	.10	.13	.15	.18	.23	.29	.35	.52
10	18	.12	.15	.18	.21	.28	.35	.42	.62
10	30	.13	.17	.21	.24	.33	.41	.49	.71
10	60	.14	.19	.23	.28	.37	.47	.56	.78
12	2	.06	.07	.07	.07	.08	.09	.10	.12
12	3	.07	.07	.08	.09	.10	.11	.13	.17
12	4	.07	.08	.09	.10	.11	.13	.15	.21
12	5	.08	.09	.10	.11	.13	.15	.18	.25
12	6	.08	.09	.10	.12	.14	.17	.20	.29
12	8	.09	.10	.12	.13	.17	.20	.24	.36
12	10	.09	.11	.13	.15	.19	.23	.28	.42
12	12	.10	.12	.14	.16	.20	.25	.30	.46
12	18	.11	.13	.16	.18	.25	.31	.37	.56
12	30	.12	.15	.18	.22	.29	.37	.44	.65
12	60	.13	.17	.21	.25	.33	.42	.51	.73
16	2	.06	.06	.06	.07	.07	.08	.09	.10
16	3	.06	.07	.07	.08	.09	.10	.10	.14
16	4	.07	.07	.08	.08	.10	.11	.12	.17
16	5	.07	.08	.08	.09	.11	.13	.14	.20
16	6	.07	.08	.09	.10	.12	.14	.16	.23
16	8	.08	.09	.10	.11	.14	.16	.19	.28
16	10	.08	.10	.11	.12	.15	.19	.22	.33
16	12	.09	.10	.12	.13	.17	.20	.24	.37
16	18	.10	.11	.13	.15	.20	.25	.30	.46
16	30	.11	.13	.15	.18	.24	.30	.37	.56
16	60	.12	.15	.18	.21	.29	.36	.44	.65

For any experimental design, the estimates of components of variance can be solved for in a similar manner. Starting with the basic error component, substitute the observed error mean square for its expected value. Successively, substitute previously obtained estimates of components of variance into higher level $E(MS)$ equations in such a way that, at each step, there is one unknown variance component to be solved for. It is entirely conceivable that, in planning a new experiment, some components of variance to be present in the new experiment may be estimated from one previous analysis and other components from another.

Estimating Component of Variance for Hypothesis. One of the most difficult problems in use of the procedure described herein is that of estimating the magnitude of the component of variance due to treatment. Three specific suggestions can be made: Results from other studies can be used to estimate the likely magnitude of true treatment differences for the criterion measure involved. For example, where new psychiatric drugs are being evaluated using standard measurement techniques, one might require that the power be sufficient to detect treatment differences comparable in magnitude to those previously observed with other active drugs. Secondly, one may actually attempt to estimate the probable range and distribution of treatment means. Where one can hypothesize the order of treatment means, the probable value of the largest and smallest mean can be estimated and other means inserted at intermediate points. The component of variance due to treatments can be estimated using Equation 5, where B_i is the deviation of the individual treatment mean about the grand mean. Finally, one can simply decide on the minimum deviation of an individual treatment mean about the grand mean which should be considered important to detect. The component of variance due to treatments can then be estimated from the minimum meaningful deviations using Equation 5. This latter procedure has been recommended by both Patnaik (1949) and Winer (1962).

$$\sigma_t^2 = \sum^{k} B_i^2/k, \qquad [5]$$

where B_i is the hypothesized deviation of the individual treatment mean, and k is the number of treatments.

The noncentrality parameter λ can be computed from the component of variance due to treatments, σ_t^2, and the components in the $E(MS)_\epsilon$ using the general form of Equation 3:

$$\lambda = km\sigma_t^2/\sigma_o^2.$$

Alternative Experiments and Cost. The components of variance associated with various error factors are empirical realities with which the investigator must contend. On the other hand, the coefficients associated with these components of variance in the $E(MS)$s are parameters of the experimental design under control of the E. The coefficient m may represent the number of independent determinations on each S. The coefficient n represents the number of Ss per group. The coefficient r may represent the number of levels of a control variable. These are all matters of experimental design.

One of the most direct ways to examine the effects of variations in the magnitudes of coefficients in the $E(MS)$s is to consider the F ratio. The estimated F ratio, formed as the ratio of two $E(MS)$s, can be considered the single value most representative of the noncentral F' distribution (Fig. 1). Although the relationship is not a linear one, an increase in the magnitude of the estimated F ratio tends to correspond to an increase in power. As we shall see in a later section, the ratio of the estimated F ratio (derived from the $E(MS)$s and components of variance) to the critical value F_α provides a convenient and useful index of relative power for comparing alternative experimental designs. The larger the estimated F ratio relative to F_α, the larger will be the area under F' and to the right of F_α.

Consider an experimental design involving m independent measures on each of n subjects in each of p randomly selected hospitals (h) for k different treatment groups. Under proper assumptions, this could be viewed as a hierarchal design having the following $E(MS)$s (Winer, 1962, p. 188):

$$E(MS_t) = \sigma_\epsilon^2 + m\sigma_s^2 + nm\sigma_h^2 + pnm\sigma_t^2$$

$$E(MS_h) = \sigma_\epsilon^2 + m\sigma_s^2 + nm\sigma_h^2$$

$$E(MS_s) = \sigma_\epsilon^2 + m\sigma_s^2$$

$$E(MS_\epsilon) = \sigma_\epsilon^2.$$

The proper F ratio for testing the significance of treatment differences would contain the following components of variance and coefficients:

$$F = \frac{\sigma_\epsilon^2 + m\sigma_s^2 + nm\sigma_h^2 + pnm\sigma_t^2}{\sigma_\epsilon^2 + m\sigma_s^2 + nm\sigma_h^2}. \quad [6]$$

This same F ratio can be expressed in alternate form by dividing both the numerator and denominator by pnm. In this form, it becomes obvious that for given values of σ_ϵ^2, σ_s^2 and σ_h^2 the estimated F ratio depends upon m, n, and p:

$$F = \frac{\dfrac{\sigma_\epsilon^2}{pnm} + \dfrac{\sigma_s^2}{pn} + \dfrac{\sigma_h^2}{p} + \sigma_t^2}{\dfrac{\sigma_\epsilon^2}{pnm} + \dfrac{\sigma_s^2}{pn} + \dfrac{\sigma_h^2}{p}}. \quad [6a]$$

An increase in any one of the coefficients m, n, or p has the effect of increasing the numerator of the estimated F relative to the denominator, but not at the same rate. Increase in m, no matter how great, can only affect F through decrease in the first term of both numerator and denominator. Increase in n tends to affect F through reduction in the first two terms in both numerator and denominator. Even if the number of *S*s in each hospital were increased without limit, the term associated with true variations among hospitals would not be affected. An increase in p tends to reduce all three terms common to both the numerator and denominator. As p increases without limit there is no limit on increase in the estimated F value. Other things equal, increase in the largest unit (hospitals instead of individuals) will effect the greatest increase in F. To what extent one type of unit should be increased as opposed to another will depend upon the relative magnitudes of the various components of variance and upon the relative costs of adding the different kinds of units.

Given the estimated F ratio of Equation 6 and estimates of the components of variance σ_ϵ^2, σ_s^2, σ_h^2, and $\sigma_t^2 = \sum B_i^2/k$, it is possible to calculate λ for any combination of m, n, and p and then to look up the approximate power of the test in Table 1.[3] Under restrictions of limited resources, it is possible to compare the power of alternative tests of the same hypothesis and to choose that allocation of resources which will yield maximum power. The noncentrality parameter λ for the hierarchal design of Equation 6 is as follows.

$$\lambda = kpnm\sigma_t^2/(\sigma_\epsilon^2 + m\sigma_s^2 + nm\sigma_h^2), \quad [7]$$

where the numerator is k times the last term in the numerator of Equation 6 and the denominator is the denominator of Equation 6.

If there were no practical limitations on available resources, the power of a test could be increased to unity by increasing experimental units represented by p, n, and m. In actuality, time and resources will not permit this. Each sampling unit (duplicate measure, individual-within-hospital, hospital, etc.) usually has a cost associated with it. Usually, the cost associated with one type of sampling unit will be different from the cost of another. It is this fact that frequently causes one allocation of resources to be superior to all others in maximizing power relative to cost.

Choice Among Permissible Experiments. A "permissible" experiment is one which can be conducted within limitations of resources available to *E*. Since power increases monotonically with increase in number of sampling units, only those possible experiments which utilize all allocated resources will be considered "permissible." For example, suppose that $2,000 were available for an experiment. The set of "permissible" experiments would include all relevant experiments costing approximately $2,000.

Suppose that a researcher has $2,000 to study the effects of a new tranquilizing drug on serum-free fatty acids (FFA) following performance under stress. It costs $40 to train each *S* up to the performance criterion. Each test trial, 1 week apart to eliminate any serial correlation, costs $10. Each laboratory determination of FFA level from a blood sample costs $2. A hierarchal design is considered in which m independent FFA determinations are to be made from the blood sample taken fol-

---

[3] The values in Table 1 were calculated by numerical integration in two parts of the noncentral F distribution function using Simpson's rule. Some 30 values over a wide range of m_1, m_2, λ, and F' have been checked against values computed using tables of the incomplete *β-function* as described by Pearson (1934) and have been found to agree to three places. Thus, the two-place values reported in this table can be accepted as accurate to the second decimal place.

TABLE 2

ANALYSIS OF VARIANCE FROM PREVIOUS STUDY

Source	*SS*	*df*	*MS*	σ^2
Subjects	1,917,947	23	83,389	5,683.5
Trials	2,730,312	72	37,921	18,809.3
Duplicates	29,036	96	302	302.5

lowing each of p independent trials for each of n randomly assigned *S*s in $k = 2$ (drug and placebo) groups. This is a three-level hierarchal design similar to the one considered in the previous section. The estimated F ratio is as follows:

$$F = \frac{\sigma_\epsilon^2 + m\sigma_t^2 + pm\sigma_s^2 + npm\sigma_a^2}{\sigma_\epsilon^2 + m\sigma_t^2 + pm\sigma_s^2}. \quad [8]$$

Following the general form of Equation 3, the noncentrality parameter λ necessary to estimate the power of the test is the ratio of k times the right-hand term in the numerator of F to the denominator of the F.

$$\lambda = knpm\sigma_a^2/(\sigma_\epsilon^2 + m\sigma_t^2 + pm\sigma_s^2), \quad [9]$$

where k is the number of treatments, n the number of *S*s per group, p the number of independent trials per *S*, m the number of independent FFA determinations per trial, and where σ_ϵ^2 is the component of variance due to differences in duplicate measures on the same trial; σ_t^2 is the component of variance due to intraindividual variability between trials, σ_s^2 is the component of variance due to true long-term differences between *S*s within groups, and $\sigma_a^2 = \sum B_i^2/k$ is the hypothesized component of variance due to true treatment differences.

A search of the literature yields estimates of the magnitudes of components of variance associated with measurement error, σ_ϵ^2, intra-individual variability across trials, σ_t^2, and with average *S* differences within groups, σ_s^2 (Hollister & Overall, 1964). From Table 2, the following values are obtained: $\sigma_\epsilon^2 = 302.5$, $\sigma_t^2 = 18{,}809.3$, $\sigma_s^2 = 5{,}683.5$. It is obvious from the summary table that the component of variance due to within-trial errors of measurement is negligible, $\sigma_\epsilon^2 = 302.5$. Only values of $m = 1$ and $m = 2$ are considered in the list of permissible experiments, since it is obvious that further increasing the number of duplicate determinations would be a waste of resources and should reduce overall power in testing treatment differences.

All experiments considered permissible for testing treatment differences within the restriction of $2,000 total expenditure are listed in Table 3. At $40 for training, $10 for a single trial, and $2 for each laboratory determination, the minimum cost per *S* is $52. This means that the *maximum* number of *S*s per group will be 19 if one is restricted to a total cost of less than $2,000. The *minimum* number of *S*s which will yield necessary within-group error estimate is 2 per group. The permissible experiments in Table 3 were chosen

TABLE 3

k	n	m	p	λ	Power	$E(\hat{F})/F_\alpha$
2	2	1	38	1.62	.12	.09
2	2	2	32	1.59	.12	.09
2	3	1	24	2.31	.21	.28
2	3	2	20	2.26	.21	.28
2	4	1	17	2.94	.29	.41
2	4	2	15	2.88	.29	.41
2	5	1	13	3.49	.38	.52
2	5	2	11	3.37	.35	.51
2	6	1	10	3.95	.43	.60
2	6	2	9	3.85	.42	.59
2	7	1	8	4.33	.47	.67
2	7	2	7	4.17	.46	.65
2	8	1	7	4.75	.51	.73
2	8	2	6	4.52	.50	.71
2	9	1	5	4.73	.52	.75
2	9	2	5	4.75	.52	.75
2	10	1	5	5.26	.56	.82
2	10	2	4	4.79	.53	.77
2	11	1	4	5.25	.57	.83
2	11	2	3	4.58	.52	.76
2	12	1	3	4.98	.55	.81
2	12	2	3	4.99	.55	.81
2	13	1	3	5.39	.58	.87
2	13	2	2	4.28	.50	.74
2	14	1	2	4.59	.53	.78
2	14	2	2	4.61	.53	.78
2	15	1	2	4.92	.55	.82
2	15	2	1	3.04	.37	.60
2	16	1	1	3.23	.39	.63
2	16	2	1	3.25	.39	.63
2	17	1	1	3.43	.40	.65
2	17	2	1	3.45	.44	.66
2	18	1	1	3.63	.46	.68
2	18	2	1	3.65	.46	.68
2	19	1	1	3.83	.48	.71

Note.—In this table, k is the number of treatment groups, n is the number of *S*s per group, m is the number of duplicate measures per *S* each trial, p is the number of independent trials per *S*, and λ is the noncentrality parameter calculated using Equation 9.

by systematically varying n between 2 and 19 and m between 1 and 2. With $k = 2$ treatment groups, the number of trials per *S* was calculated as the integer portion of

$$p = (2000 - 40nk)/(10nk + 2nmk).$$

Consideration of expected elevation in serum-free fatty acid levels suggested that a difference of 100 units between drug and placebo would be meaningful and important to detect. The component of variance due to treatment factors was computed using $B_1 = B_2 = 100/2$:

$$\sigma_a^2 = (50^2 + 50^2)/2 = 2500.$$

From values m, n, p, and k and from estimates $\sigma_\epsilon^2 = 302.5$, $\sigma_t^2 = 18{,}809.3$, $\sigma_s^2 = 5{,}683.5$, and $\sigma_a^2 = 2{,}500$, the values of noncentrality parameter λ were computed for each permissible experiment by use of Equation 9. In this example, the *df* are $N_1 = k - 1$ and $N_2 = 2(n - 1)$. Entering Table 1 with N_1, N_2, and λ, an estimate of the power of each permissible allocation of resources was obtained. Under the conditions imposed by costs of training, testing, and laboratory determinations, and in view of the relative magnitudes of components of variance associated with various sources of error, the most powerful test of treatment differences in this example would appear to involve 13 *S*s per group with three independent trials each and only one laboratory determination per trial.

In Figure 2 are presented graphs of the power for single and duplicate FFA determinations by number of trials per *S* with number of *S*s determined by the restriction that total cost should not exceed \$2,000. It is apparent that variations in the allocation of resources produce marked differences in power of the test of treatment differences. In this example, as in many practical problems, the differential cost of various sampling units renders one particular allocation of resources most powerful. It is also interesting to note that, because of the high reliability of the measurement technique within trials, the allocation of resources to two determinations per trial ($m = 2$) is consistently less powerful than corresponding allocations of resources such that $m = 1$. Had the simple error of measurement been greater, duplication of lab determinations at \$2 apiece might have proven advantageous.

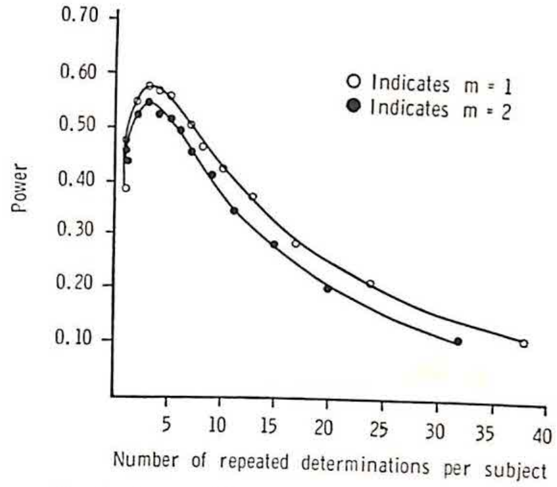


FIG. 2. Power of *F* test for various numbers of trials per *S*.

The Zero-Overhead Case. An important question in psychological research is concerned with whether or not to test each *S* more than once in order to increase the reliability of measuring some highly variable performance. If there is a fixed cost associated with each test, but no difference in cost associated with whether the testing involves the same or different *S*s, then the answer is that it is unquestionably better to use more *S*s and test each one only once. *This is true no matter how unreliable the measurement involved.*

Consider an investigator interested in evaluating the effect of a new tranquilizing drug on serum FFA level, as in the former example, except that there is no separate cost for training. Each test trial costs \$10, whether on the same or different *S*. Since it is recognized that laboratory errors are negligible, no duplicate determinations from each trial are considered. Assuming independence and normality within *S*s, a simple hierarchal design is appropriate:

$$F = \frac{\sigma_\epsilon^2 + m\sigma_s^2 + nm\sigma_a^2}{\sigma_\epsilon^2 + m\sigma_s^2}.$$

The noncentrality parameter λ can be computed as follows for testing treatment effects:

$$\lambda = knm\sigma_a^2/(\sigma_\epsilon^2 + m\sigma_s^2).$$

Suppose that \$800 were available for the research. At a minimum cost of \$10 per *S*, the

TABLE 4

k	n	m	p	λ	Power
2	2	1	20	1.51	.12
2	2	1	18	1.48	.12
2	2	1	16	1.45	.12
2	2	1	14	1.42	.12
2	3	1	12	2.06	.20
2	4	1	10	2.63	.27
2	5	1	8	3.10	.33
2	6	1	6	3.38	.36
2	8	1	5	4.21	.47
2	10	1	4	4.78	.53
2	13	1	3	5.39	.58
2	20	1	2	6.56	.71
2	40	1	1	8.07	.82

Note.—In this table, k is the number of treatment groups, n is the number of *S*s per group, m is the number of duplicate measures per *S* each trial, p is the number of independent trials per *S* and λ is the noncentrality parameter calculated using Equation 9.

maximum number of *S*s per group (drug and placebo) would be 40. The *minimum* number of *S*s per group would be 2, with 20 trials per *S*. The set of permissible experiments includes all combinations of m and n between these extreme values such that total cost is held to \$800. These permissible experiments are listed in Table 4. The general conclusion is that, if there is no additional special cost associated with the use of different *S* for each determination, the most powerful test results from using a maximum number of *S*s with only one determination each. This conclusion is valid no matter how unreliable the measurement involved.

Use of $\hat{F}$ *in Estimating Power.* The power of an F test depends upon the proportion of the F' distribution lying to the right of F_α (see Fig. 1). While integration of the noncentral F' distribution provides a problem for the researcher without a computer handy, it is obvious that a central estimated value in the F' distribution provides a good indication of the general location of that distribution. We will define the estimated $\hat{F}$ ratio as the ratio of $E(MS)$s for treatment and error. The farther that the estimated $\hat{F}$ ratio exceeds F_α, the larger will tend to be the proportion of the F' distribution which will be found to lie beyond F_α. Thus, the ratio $E(\hat{F})/F_\alpha$ provides an appropriate index of the power of any F test.

The $\hat{F}$ ratio can be estimated from the $E(MS)$s, and estimates of components of variance included in them:

$$\hat{F} = \frac{\sigma_\epsilon^2 + m\sigma_t^2 + pm\sigma_s^2 + npm\sigma_a^2}{\sigma_\epsilon^2 + m\sigma_t^2 + pm\sigma_s^2}. \quad [10]$$

The estimated $\hat{F}$ ratio can be written from $E(MS)$ values tabled for most experimental designs. In the example considered in a previous section, estimates of $\sigma_\epsilon^2 = 302.5$, $\sigma_t^2 = 18{,}809.3$, $\sigma_s^2 = 5{,}683.5$, and $\sigma_a^2 = 2{,}500$ were obtained from the literature. The power of tests representing different allocations of the resources are presented in Table 3. For the same set of permissible experiments defined in terms of m, n, and p, the values of estimated $\hat{F}$ ratios were computed by substituting into Equation 10. The ratio of $\hat{F}$ to F_α for specified df was computed for each permissible experiment, and the results are entered in the right-hand column of Table 3. A scatter plot of the relationship between calculated true power and $\hat{F}/F_\alpha$ ratio is shown in Figure 3. The linear correlation between power of test calculated as integral of noncentral F' distribution and *relative* power as estimated by $\hat{F}/F_\alpha$ was found to be .988 in this example. These results suggest the usefulness of the $\hat{F}/F_\alpha$ approximation as a basis for comparing alternative permissible experiments. As can be seen by inspection of the right-hand column of Table 2, this approximate procedure would have resulted in

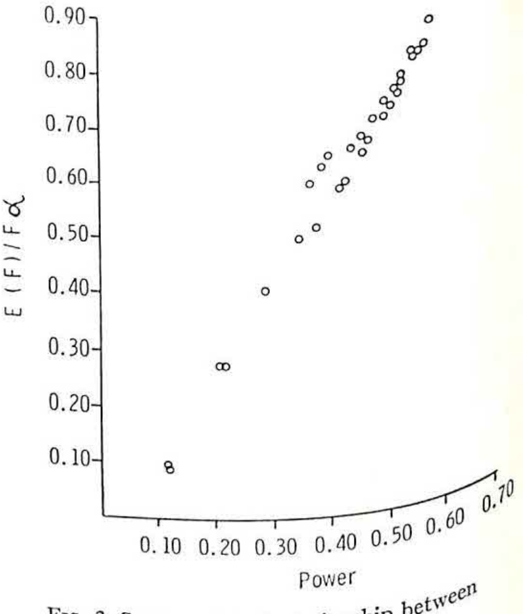


FIG. 3. Scatter plot of relationship between power and $E(\hat{F})/F_\alpha$ index.

selection of an experiment involving 13 *S*s per group with three independent test trials each. The end result is identical to that obtained by calculation of exact power values. In cases where *df* or λ values are not adequately approached in the tables, the $\hat{F}/F_\alpha$ approximation should provide an adequate basis for comparing alternative permissible experiments. Actually, for purposes of designing experiments, rather crude approximation in estimating relative power should provide a useful guideline for avoiding unfortunate allocation of resources and for choosing an advantageous combination.

In Figure 4 are presented graphs of the $\hat{F}/F_\alpha$ ratio plotted against number of trials per *S*. These curves can be compared directly with the power curves presented in Figure 2 for the same data. The curves peak at three trials per *S*, and the consistent superiority of a single determination per trial in this particular example is apparent. Comparison of Figures 2 and 4 makes obvious the utility of the simple $\hat{F}/F_\alpha$ ratio where tabled power values are not available.

Derived Formulas for Optimal Design. Several statisticians have worked on the problem of allocation of resources in an attempt to find a simple direct way to calculate the optimal number of each type of experimental unit. As far as we can find, others have not actually used the logical, but complex, criterion of *power* as a basis for optimizing. One approach which may yield results highly similar to those based upon maximum power is that of minimizing the total error variance (Brooks, 1955). For a given true treatment difference, reduction in error will increase the estimated *F* ratio.

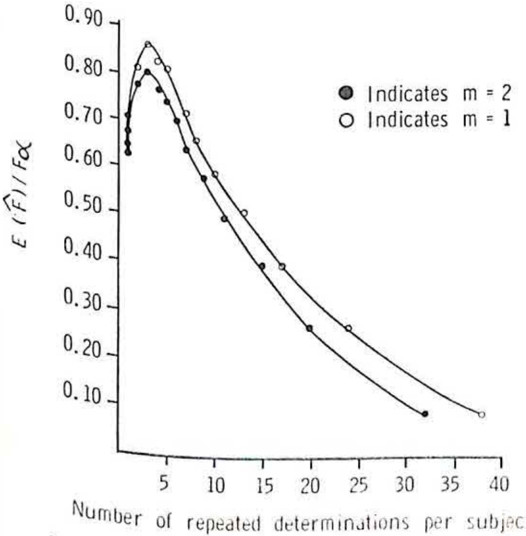


FIG. 4. $E(\hat{F})/F_\alpha$ index for various numbers of trials per *S*.

$$F = \frac{\sigma_\epsilon^2 + m\sigma_t^2 + pm\sigma_s^2 + npm\sigma_a^2}{\sigma_\epsilon^2 + m\sigma_t^2 + pm\sigma_s^2}$$

The total cost of the experiment can be expressed in terms of the costs of the different kinds of experimental units:

$$C = nC_1 + pnC_2 + mpnC_3$$

Using methods of calculus, it is possible to solve for values of *p*, *m*, and *n* which will minimize the denominator of *F* relative to cost. It will be noted that the results from this approach do not take into consideration the magnitude of the hypothesized true treatment differences, *df*, or alpha level. Hence, the minimizing of error variance alone will not always result in the most powerful test. It is, nevertheless, a useful approach with which the psychologist should become familiar.

We have attempted an analytic solution for the problem of maximizing the ratio of estimated $\hat{F}$ to F_α subject to fixed cost. The Lagrange method was used and a solution was found for the problem represented in Table 3. We have concluded that the same solution can be obtained more easily through tables to evaluate power of alternative permissible experiments, or through comparison of $\hat{F}/F_\alpha$ ratios for alternative permissible experiments. Construction of a table of alternative experiments, such as Table 3, has the advantage of displaying changes in power for comparison among various practical alternatives. In some cases, reasons other than economic may suggest the use of a combination which is not the single most powerful. Through the tabling of alternative permissible experiments, a *good* alternative can be chosen.

REFERENCES

BROOKS, S. The estimation of an optimum subsampling number. *Journal of the American Statistical Association*, 1955, **50**, 398–415.

FEDDERER, W. T. *Experimental design*, New York: Macmillan, 1955.

FELDT, L. S., & MAHMUD, M. W. Power function charts for specification of sample size in analysis of variance. *Psychometrika*, 1958, **23**, 201–210.

FISHER, R. A. The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society, Series A*, 1928, **121**, 654.

GAITO, J. Repeated measurement designs and counter balancing. *Psychological Bulletin*, 1961, **58**, 46–54.

HOLLISTER, L. E., & OVERALL, J. E. Sources of variation in repeated measurement of plasma free fatty acids in rested fasting subjects. *American Journal of Clinical Nutrition*. 1964, **15**, 149–157.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.

PATNAIK, P. B. The non-central chi-square and F-distributions and their applications. *Biometrika*, 1949, **36**, 220–232.

PEARSON, E. S., & HARTLEY, H. O. Charts of the power function for analysis of variance tests, derived from the non-central F-distribution. *Biometrika*, 1951, **38**, 112–130.

PEARSON, KARL. *Tables of the incomplete beta function*. Cambridge, Eng.: Cambridge Univer. Press, 1934.

SATTERTHWAITE, F. E. Synthesis of variance. *Psychometrika*, 1941, **6**, 309–316.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received January 22, 1965)

Psychological Bulletin
1965, Vol. 64, No. 5, 351–364

DEFINITIONS AND METHODOLOGY IN THE STUDY OF AGGRESSION[1]

HARRY KAUFMANN
University of Toronto

The literature on aggression and hostility is viewed in the light of 2 major criticisms: (a) Definitions of terms with obvious denotations of hypothetical constructions are attempted in purely operational terms, (b) The phenotype of aggression has impaired a sound investigation of the many genotypic determinants of aggressive behavior. Present trends of studying aggression as an end result of a variety of stimulus, dispositional, and reward variables are considered, and some amplifications are suggested.

The issue of aggressive tendencies and aggressive behavior has long been a source of particular interest to psychologists and especially in the last few years has produced several exhaustive and insightful treatments (Berkowitz, 1958, 1962, 1964; Buss, 1961; McNeil, 1959; Pepitone, 1964). Nevertheless, in spite of the great need for, and interest in, a more complete understanding of the determinants of aggression, we are still on quite uncertain ground as far as relevant theoretical interpretations are concerned.

Frustration-Aggression Hypothesis

The original frustration-aggression (F-A) hypothesis (Dollard, Doob, Miller, Mowrer, & Sears, 1939) defined frustration as "that condition which exists when a goal-response suffers interference [p. 11]," and aggression as "an act whose goal response is injury to an organism (or organism-surrogate) [p. 11]." The hypothesis then asserted a one-to-one relation between frustration and aggression, although, and this is often overlooked in later criticisms, the authors did not argue that aggression in such cases is always observable. N. E. Miller (1941) rephrased the presumed relationship in such a manner as to admit, in place of the heretofore operationally definable aggression, an "instigation to aggression" which can be conceived as being either so weak or so low in response hierarchy as to be undetectable. Brown and Farber (1951), in turn, modified the definition of frustration from its original meaning of "interference with a goal directed activity" to a hypothetical, internal arousal response to certain stimulus situations. This response, apparently largely autonomic in nature, could then serve as a mediating stimulus for aggressive responses. The objective definition of frustration has been questioned repeatedly (see Maslow, 1941; Pastore, 1952; Rosenzweig, 1944). Among more recent writers, Berkowitz (1962) and McNeil (1959) continued to lean toward the F-A hypothesis. McNeil (1959) continued to view aggression within the framework of the F-A hypothesis, though with some modifications. He drew freely upon psychoanalytic concepts and learning principles, but his overview aimed more at description than at theoretical analysis. In his recent book, Berkowitz (1962, p. 30) first noted the failure to define frustration properly, then (p. 62) approvingly cited McDougall (1926) and viewed having one's toe stepped on as a frustration of the impulse to self-assertion. He also noted (p. 138) that aggression need not always be the dominant response to frustration but may be "inhibited" under certain conditions.

One is thus forced to conclude that the conceptual value of even the modified versions of the hypothesis advocated by McNeil and by Berkowitz is no longer discernible, since both its independent and dependent variables are left so protean as to be refractory to operational definition, a sure sign that, with due recognition of its influential role in experimental research, it has outlived its usefulness. This conclusion is shared by such diverse sources as Seward (1945a), Menninger (1942, p. 296), and Buss (1961), who arrived at it by the different routes of animal behavior, psychiatry, and social psychology.

---

[1] The present paper was completed with the help of a National Science Foundation Research Grant and a Sigma Xi Grant-in-Aid.

Physiological and Genetic Determinants of Aggressive Behavior

The attempts to subsume aggression under one global theoretical edifice have assumed other guises. Experiments by Beach (1945) and Beeman (1947) attested to a direct relationship between hormonal balance and fighting behavior, but Scott (1958) reported successfully training mice not to fight and, more generally, produced convincing evidence against the so-called "fighting instinct" in mammals. Bevan, Daves, and Levy (1960) indicated that testosterone concentration is less effective in producing fighting behavior in rats than previous fighting experience, even though the hormonal composition appears to weigh heavily in the initial shaping of response patterns. Clark and Birch (1945) noted that physiological explanations are of little value in accounting for complex human behavior. Scott and Fredericson (1951) extensively documented their argument that hormonal factors are irrelevant in fighting behavior subsequent to its acquisition. Even Tinbergen (1953) found no grounds for postulating a general fighting instinct among animals.

We can therefore concur with McNeil (1959, p. 198) in the well-founded conclusion that the explanation of aggression in humans is not furthered by the invocation of physiological or genetic factors.

Redefinition of Aggression

The term aggression has not always been used unequivocally. Even if we disregard the distinction between aggression and "aggressive behavior" and use both terms synonymously to denote only overt behavior, several points remain unclear. First of all, traditional definitions of aggression as a response having for its goal the injury of a living object are not necessarily applicable to so-called instrumental aggression, or behaviors aimed at removing obstacles between the aggressor and a goal, where injury may result only incidentally. Ambiguity exists within the narrower traditional definition. Buss (1961), for instance, used a somewhat similar definition —"a response that deliver noxious stimuli to another organism [p. 3]"—but soon ran into a contradiction when (p. 32) he considered Gellerman's (1956) word association task and his own scrambled sentences task as measures of aggression, though both tasks are clearly "intransitive" in that they have no direct target. Thus, these devices measure what might more appropriately be classified as "potential aggression" or "aggressive tendency," which we might define as the probability for a given individual to emit an aggressive response in a given situation. A similar question might be raised with regard to various well-established fantasy measures of aggression, where noxious stimuli are delivered in a purely autistic manner in that they not only do not result in actual injury to the object (nor in its removal, where obstruction-removing aggression is concerned), but presumably have no initial subjective probability of reaching the object. Questionnaires, which aim at assessing a person's opinions about another individual are equally equivocal in their meaning. Is it justifiable to classify as aggression an opinion which, in terms of the respondent's perception, has no chance of affecting the target of his remarks? The preoccupation with violent materials measured by, say, a sentence completion task, may, but does not necessarily denote aggressive tendencies, that is, an increased likelihood of committing aggression in the presence of a suitable object. Buss classified thwarted or uncompleted aggressive acts as aggression, but it might be argued that this classification does not account for acts which patently never could have reached their object. It might be of value to differentiate, then, whether the term aggression is used to denote the delivery of a noxious stimulus which, whether completed or not, initially has a subjective probability >0 of reaching its object, or whether simple emission of certain behaviors by the aggressor is sufficient. At first sight, it seems possible to view aggression as a pure response, similar to eating or running, and measure its intensity. This is the view proposed by Buss (p. 3 ff.) who decried the inclusion of "intent" in his quest for a purely behavioral definition. But here too problems of definition arise: When does the sudden extension of the forearm have the charactersitics of aggression as opposed to, say, physical exercise? And can mere verbal production of a class of words classified as aggressive be taken as indicative of aggression, unless inferences about intent and expectations are made? Feshbach (1964) also was careful to separate the mere expression of anger (and, presumably, derogation) from aggressive behavior. The facilitation of aggression by anger, on the other hand, is readily explained by conditioning principles.

Little attention has also been paid to the question of the victim's awareness that he is the object of an aggressive act. Where the goal of the aggressive act is injury to the victim and such an act is completed successfully according to aggressor's evaluation, is such injury rewarding even if the victim does not perceive the source of his injury, or perhaps does not even realize that he has been injured? A problem which follows directly from the issue of effectiveness of aggression and awareness of the victim (as perceived by the aggressor) is that of measuring the intensity of the aggressive act. The most obvious indicator of intensity is the strength of the response emitted by the aggressor. But where the aggressor perceives two or more objects of aggression as differentially susceptible and, wishing to obtain equivalent effects, gauges his aggression accordingly, it may be misleading to maintain that he was aggressing more violently against a more powerful than a less powerful object.

Finally, the definition of aggression is unclear where an aggressive act is committed reluctantly, under orders, or perhaps a sense of duty. Granted that much malice and injury is rationalized by the attacker as just punishment or an act of obligatory adherence to social norms, it would be begging the question to assert that aggression is never committed without some satisfaction to the aggressor arising out of injury to the victim and, conversely, that the interruption of such planned aggression is always noxious. Presumably, Abraham welcomed the interruption of his aggression against Isaac. Aggression, then, has been implicitly or explicitly defined in a variety of ways, none of which encompass all others. Feshbach (1964) made a clear and telling point for the inclusion of motivational factors in the definition of aggression. Much of the difficulty was mirrored by the attempts by Buss and others to assign behavioral definition to what is, in fact, a hypothetical construct. In essence, then, no behavioral act as such can consistently be called aggression, unless inferences about intention or expectation regarding its outcome are made. In the body of this paper, the term aggression will be used to denote behaviors which: (*a*) are transitive, that is, are directed against some object; (*b*) have a subjective probability > 0—on the part of the attacker—of reaching that object, and either removing it from attacker's goal path, or imparting a noxious stimulus to it, or both.

This definition includes even totally unsuccessful responses (provided the attacker had some initial expectation of success) but does not include certain opinions or propensities as measured by questionnaire or projective devices, except where evidence or reasonable assumption indicates that respondent expects some injurious effect to result from his actions.

Some Recent Theoretical Statements

Buss (1961, p. 198 ff.) attempted a behavioral approach to aggressiveness and defined the latter as the habit of attacking. He listed four determinants of aggressive habit strength, and defined them as follows:

1. *Antecedents of aggression.* The number of times the subject has been angered, attacked, or annoyed.

2. *Reinforcement history.* Degree and frequency of reinforcement or punishment.

3. *Social facilitation.* Cultural and subcultural norms regarding aggression.

4. *Temperament.* Such personality variables as impulsiveness, intensity of reaction, activity level, and independence (tendency toward self-sufficiency and resistance to group pressures).

However, Buss offered little in the way of a consistent theoretical reformulation, though in many respects he appeared to adopt Hullian concepts. He rejected, it is true, anger (p. 13) and frustration (p. 20) as the specific drives underlying aggression, but his use of "tendency" bears strong resemblance to "drive." Also, he spoke of "a sharp drop in anger level" as "reinforcing" (p. 199). Since his first determinant, antecedents of aggression, was apparently distinct from the second, reinforcement history, what appeared to be meant by the former was the total number of "trials," that is to say, the number of times an individual has been aroused. This variable appears related to Estes' (1950) and Guthrie's (1952) contiguity principle of learning, or it may belong more properly in the category of learned drives. The second determinant is synonymous with habit as defined by Hull, namely, a function of the number of reinforced trials.

Buss' use of this third determinant, social facilitation, was similarly ambiguous. If customary social rewards of aggression were intended, then this variable was identical to the reinforcement variable with social approval as a secondary reinforcer. On the other hand, if the degree of facilitation or inhibition of a specific aggressive act was meant, it would have been more proper to consider it an incentive variable, analogous to Hull's *K*.

An insightful and rigorous treatment was offered by Pepitone (1964). As did Buss (1961), he rejected the circularity of the F-A hypothesis and pointed to the importance of an evaluation of the stimulus. Pepitone, however, went beyond Buss and examined the cognitive implications of a potentially agression-producing situation. He began with an analysis of the "threat" stimulus postulated earlier by Maslow (1941) and Rosenzweig (1944). Such threat may be subdivided into the thwarting of various needs such as status or affiliation, absence of cognitive balance or congruence, or, finally, social pressures toward uniformity.

Pepitone concluded that even this finer grained analysis does not preclude equivocal interpretations of extant experimental data, and proceeded to report a series of experiments in which such specific potential aggression inducers as boastfulness and threat against the subject's self-evaluation on the part of an instigator were studied in interaction with that person's perceived actual worth. By demonstrating the effects of such variables as cognitive validation and self-evaluation conflict, Pepitone avoided the pitfall of a global interpretation. While his analysis deliberately confined itself to a narrow range of inquiry, it laid some important groundwork for a precise assessment of the several variables which determine the occurrence of aggressive acts. Feshbach's (1964) most recent statement came to grips with some of the difficulties which have beset studies of aggression and went a long way toward resolving them. He distinguished between incidental, instrumental, and hostile aggressive acts, thus clearly reintroducing motivation and specifically inferred intention into the paradigm. He also noted how initially aimless acts of a motor activity may become reinforced by manifestations of pain or injury, and thus develop into aggressive habits. While he distinguished such "expressive" aggression from instrumental and hostile aggression, this category could logically be dispensed with. When injury is the accidental outcome of some expressive motor act, it is not aggression; when it recurs as a function of reinforcement, either resulting in some unexpected rewards or by observation of pain, it falls into one of the two other categories.

We might go one step further and argue that even the remaining dichotomization of aggression into instrumental and hostile introduces artificial categories, based only upon observation of ultimate outcomes. Feshbach's (1964) own observation, that the elicitation of pain cues as a reinforcer of aggression seems to be restricted to humans, appears to be at direct variance with his three-factor theory of aggression. There is no a priori or parsimonious reason why hostility-engendered, injury-directed aggression should be viewed as being different from ordinary instrumental aggression. What may differ are simply the reinforcing stimuli which make such behavior more or less likely in the future.

The present argument is more generally applicable to most previous analyses of aggression, where superficial—though socially important—similarities in response patterns led to unwarranted inferences about their antecedents.

The pitfalls of such phenotypic reasoning (not at all a recent observation, by the way) become obvious if we look at such relatively simple behavior as eating. Leaving aside hypothalamic lesions, we may still ask whether the organism is eating because it has been deprived of food generally, or perhaps just vitamin C or salt, or whether it eats to avoid shock, and so on ad libitum.

"Hostility" as a Theoretical Construct

A large number of experiments deal with the differential effects of instigation upon "high-hostile" versus "low-hostile" individuals. Hostility is usually defined in a variety of ways, such as scores on one or another of the current hostility questionnaires, which frequently show only moderate correlation, responses of projective or completion devices, and judgments by other persons. This global notion of hostility was questioned by Buss (1962). Earlier studies by Buss and his associates (Buss & Durkee, 1957; Buss, Durkee, & Baer, 1956), attempted to partition responses on an inventory into various categories such as suspicion, indirect hostility, etc., and extracted two factors roughly equivalent to hostility, a tendency to view the environment as inimical, and aggression, the tendency to inflict injury.

Buss' eminently useful approach of first drawing a clear distinction between aggression and hostility and then factorizing hostility into several components was vitiated by his later comparison of hostility with aggression and by the assertion (p. 204) that hostility is more specific and focused than aggression. In terms of this partition, the selective racist would be considered as hostile, but the sociopath, who aggresses indiscriminately, would not. In all, one would surmise that Buss' formulation should be able to dispense also with the construct of hostility. If, consistent with a Hullian formulation, we simply assume that aggression has been rewarded more frequently with regard to some stimuli than to others, then no postulate of differential personality characteristics is required, and differential aggression is adequately explained in terms of stimulus, habit strength, incentive value, inhibition, and a discrimination gradient.

Somewhat surprisingly, as measured by questionnaire, under certain conditions hostility may be a negative predictor of the particular kind of aggression which might be inferred from punitivity toward an offender. Kaufmann and Marcus (1965, in press) found that high-hostile college subjects displayed greater leniency toward a juvenile delinquent described as a college student than low-hostile subjects, but were more punitive than low-hostile subjects when the same offences were ascribed to a door-to-door salesman, thus bearing out some correlational findings by Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) and experimental results by Steiner and Johnson (1963). It is also a common observation that highly hostile individuals sometimes contract allegiances with ties of intense personal loyalty toward other, perhaps similarly hostile, individuals. On the other hand, the tolerant liberal is often quite hostile toward the intolerant, jingoistic know-nothing. It is not clear whether what we mean by describing an individual as hostile is that he tends to respond aggressively toward individuals or groups possessing specific stimulus characteristics, to perceive a specific kind of response as socially desirable, or even to express his aggressive feelings with greater freedom than a so-called low-hostile person. It is not maintained that hostile, extrapunitive, ethnocentric, or high F-scoring individuals do not exhibit some behaviors which set them apart from their counterparts along these dimensions. But it is highly doubtful that we can rely upon these classifications to predict targets and intensities of aggressive behavior, irrespective of other variables.

Emotional Arousal or "Anger"

Much confusion has prevailed with regard to the determinants and properties of anger. The observation that aggression is usually accompanied by anger makes all the more necessary a thorough study of possible causative, rather than correlational phenomena. It is usually taken for granted that anger at least precedes aggression and, most likely, causes or at least facilitates it. The physiological nature of anger and the voluntary one of aggression have, for instance, led to the assumption that anger is classically conditioned to certain stimulus situations and is, therefore, largely outside the control of the person. Thus, Buss (1961, p. 102) pointed to the preparatory role of physiological arousal toward subsequent aggression. The observation, by Sears, Maccoby, and Levin (1957), that anger appears to be an instinctive response may be of value where infants are concerned but, as all explanations leaning upon the concept of instinct for explanation, is of little value in understanding and predicting complex human behavior.

We are handicapped once more by the absence of a unified definition. Berkowitz (1964) spoke of anger in the absence of any aggressive behavior tendency. Buss (1961) clearly distinguished between the emotional aspects of arousal and the voluntary ones of aggression, while Schachter and Singer (1962) presented impressive evidence that anger may be largely a cognitive labeling of physiological arousal. Anger, though, is a term used to denote a multiplicity of phenomena, which are not necessarily related.

1. Anger has been defined as the diffuse, maladaptive agitation or rage response displayed by lower organisms and poorly organized humans upon being presented with an unsolvable situation. This anger response involves no evaluative or instrumental components, and the aggression to which it may lead is usually quite nonspecific, perhaps even a chance component of random behavior. The response patterns observed by Janis (1951), Kardiner and Spiegel (1947), and Berkowitz (1962), as well as temper tantrums (at the early, noninstrumental stage), fall within this classification. Feshbach (1964) spoke of this undifferentiated emotional arousal as an instigation to hit rather than to hurt. However, the consummatory nature of the hitting response is open to question.[2]

2. Anger recurs as a result of classical conditioning. Initially unlabeled emotional arousal has in the past of an individual's experience been conditioned to stimulus situations regardless of subsequent responses. When such a stimulus situation recurs, so should the arousal, which is then labeled anger in the manner described by Schachter and Singer (1962). Since labeling has the property of providing gratifying cognitive clarity, it may on subsequent occasions be mediated in the absence of the "model."

3. The third possibility, one that heretofore apparently has been largely ignored, is that of anger conditioned classically to certain preparatory responses which may initially have been selected independently of emotional arousal. The possibility of anger as a classically conditioned response should not, however, be interpreted as a necessarily permanent pairing of certain stimulus situations and emotional arousal. Recent findings by Shapiro, Crider, and Tursky (1964) and in Liu's (1964) two-factor theory of classical conditioning suggest that anger (or emotional arousal in general) is not a built-in, unalterable response to certain stimuli, but becomes established only if its presence fulfils an instrumental role.

4. Anger accompanying preparatory aggressive responses may also be initially and deliberately instrumental. Where aggression has been rewarded in the past, the subject may well have learned that emotional arousal, labeled anger, facilitated or intensified such aggression; he therefore reproduces it—works himself into a rage—thus energizing his present aggression. This susceptibility of autonomic state to cognitive influences received some support from recent studies by Brehm, Back, and Bogdonoff (1964).

5. Again a rarely considered possibility, anger may follow aggression—or increase after aggression—as a gratifying, guilt-reducing response. The subject has aggressed, feels guilty, and rationalizes by producing emotional arousal, perhaps claiming that he is still quite upset. More generally, self-stimulation to anger following aggression may have dissonance or anxiety-reducing properties.

Berkowitz (1962, p. 160) combined the pure drive notions of classical F-A theory and the "pure stimulus" theory of Zawadski (1948) and in a more recent paper (1964) saw aggression as being evoked by a "target with appropriate stimulus qualities either because the subject is angry or because particular stimuli have acquired cue value for aggressive responses from him." He distinguished between anger-produced and habitual aggression but continued to place heavy emphasis upon a preexisting state of internal arousal, despite rather ample evidence that physiological factors play an unpredictable role in aggressive behavior even in animals (Beach, 1945; McNeil, 1959; Scott, 1958)—excepting, as noted earlier, the actual acquisition of such behavior in the repertoire of some of the simpler organisms (Beeman, 1947). We have, of course, a considerable body of experimental literature which shows interesting relationships between physiological arousal and the occurrence of aggression (see Hokanson, 1961). The well-known study of Schachter and Singer (1962), showing the effects on a model upon the labeling of arousal, has already been cited. What is absent, however, is the determination of reliable antecedent-consequent relations between such arousal and aggressive behavior.

---

[2] It is also not clear whether Feshbach's observation that restraint in infancy produces a desire to hit (rather than to hurt) does not more simply represent the kind of brusque motor activity which has previously freed the infant from other, possibly inanimate, restraints.

Berkowitz himself cited ample evidence that physiological arousal alone is not sufficient to produce aggressive behavior where suitable cue stimuli are absent, and Pepitone (1964, p. 81), by implication, considered aggression without anger as a function of attempts at cognitive validation. Evidence for the occurrence of aggression without arousal ranges from Scott's (1958) mice, which, accustomed to successful aggression, displayed no arousal symptoms preparatory to aggression, to the mass executioners of an extermination camp, who (Höss, 1963) managed to commit inconceivable acts of horror without apparent displays of anger or other emotions. Emotional bluntedness or stultification in the latter case and in others similar to it would seem to be an inadequate interpretation, since many of these individuals displayed a normal range of emotions, from joy to rage, in dealing with their environment other than the camp situation. One might well venture, then, the arduous step of relinquishing the constructs of frustration and anger as either necessary or sufficient conditions for aggressive behavior. Such a position does not, of course, deny the occasional, or even usual, presence of physiological arousal where aggressive behavior occurs. It simply maintains that, to paraphrase Brown and Farber (1951), a nonemotional theory of aggression may be sufficient to account for the class of behaviors which are perhaps most economically categorized as a subset of learned social interactions. In this formulation, emotional arousal covaries highly with actual or fractional anticipatory aggressive responses, but neither determines nor is determined by them. Where present, this stage of arousal, in turn, is perceived by the person, related to other stimulus elements in the situation, and appropriately labeled in accordance with past learning. If anger is viewed as conditioned, not only to consummatory aggression but to fractional antedating goal responses, then the fact that an aggressive response seemingly becomes more likely in the presence of anger no longer implies causative or even facilitating effects of the latter.

The Notion of Displacement

The F-A group's original concept of displacement implied simply that an object other than the frustrater became the target of (undeserved) aggression. Underlying the process were two models: one in which the frustrater was feared and aggression was directed against a safer target, either a minority group (N. E. Miller & Bugelski, 1948) or, as in the well-known parable of the satanic and powerful boss and the angelic, defenseless wife, against a handy, equally safe object. Buss (1961, p. 61) named this phenomenon displacement of target and ascribed it to a lower threshold, permitting aggression to be elicited by a wider variety of stimuli. The other model proposed by the F-A group is exemplified by Miller's (1948) well-known study in which rats displaced striking behavior from white rats onto white dolls when the former became unavailable. Buss referred to this process as stimulus displacement and ascribed it, as did Miller, to a stimulus-generalization gradient.

It should be noted that the two models are not clear, nor do they account for a great majority of aggressive behavior directed against seemingly nonfrustrating targets. The often discussed phenomenon of "scapegoating" seems to follow from the model of displacement of target, but makes no prediction about which of the innumerable safe targets, including newly borns in perambulators, will finally be selected. Berkowitz (1962, p. 153) emphasized both safety and ethical aspects in the selection of a scapegoat, but the criterion of safety is directly opposed to that proposed by White and Lippitt (1960), who emphasized the self-esteem restoring aspects of aggression. Self-esteem, they argued, is restored only if a person can take some comfort in his deed. The stimulus-generalization model is unapplicable, since no plausible a priori argument has yet been presented which establishes similarity between a frustrater and a scapegoat. Buss' argument (1961, p. 247) that scapegoating requires anger as a necessary condition was also contradicted by his own definitions, earlier established, because it excluded the habitually hostile response of the prejudiced individual against the customary target, not necessarily in the presence of physiological arousal. Buss later (p. 258) did suggest that prejudice is most economically conceptualized as a learned habit, but failed to integrate this habit or response readiness with aggression. The argument of social learning and imitation was further explored by Bandura and Walters (1963). Bindra's (1959) formulation viewed displacement as the outcome of enduring response habits (similar, apparently, to Buss's hostility) which can result in undiscriminating aggression against a wide variety of targets. Berkowitz (1962, p. 135) distinguished between both culturally learned hostility and culturally enforced, hostile behavior toward members of minority groups and the displacement upon a scapegoat arising from some immediate frustration.

It is clear that both the target displacement and the stimulus displacement notions can easily become circular. A displaced target was attacked because safe and, having been attacked, must have been safe. An ultimate object of aggression is similar along some undefined generalization gradient to an original annoyer and is attacked; conversely, having been attacked, it must have been somehow similar to him. Literature and one's own experience indicates that a disliked individual can, given proper provocation, acquire a surprising resemblance to a loathsome animal, but the relationship of such an occasional metamorphosis to aggression is purely speculative.

The experimental literature in some instances supports a scapegoat hypothesis, that is to say, selective punitivity against specific targets (Weatherley, 1961), in others the data indicate a general nonspecific heightening of aggressive tendencies (Berkowitz, 1959; Lesser, 1958; Lindzey, 1950). Stagner and Congdon's (1955) data failed to support either of these hypotheses. The studies by Miller and Bugelski (1948) and Cowen, Landes, and Schaet (1959), though often cited in support of scapegoat theories, cannot conclusively be interpreted as indicating a selective heightening of aggressive tendencies since attitudes towards the subjects' peers were not measured, which makes it conceivable that antagonism against them was also higher. Berkowitz (1960, 1961) presented evidence that for highly prejudiced subjects, discriminations between thwarting- and nonthwarting-stimulus persons breaks down. Lesser (1958) also hypothesized that displacement phenomena are largely due to greater general extrapunitivity. To sum up, aggression or aggressive tendencies against a presumably nonoffending target have been viewed in terms of their observable outcome, but the understanding of the entire process has suffered from efforts, originating from the emphasis placed upon the aggressive outcome, to group widely different determinants under the same arbitrary classification.

Meanings of Catharsis

The term "catharsis," in relation to the F-A hypothesis, was originally used to denote an aggressive act followed by lowered subsequent aggression. This lowering of aggressive drive or, in terms of Miller's (1941) emendation, of instigation to aggression was ascribed to a draining away of drive leading to aggression. Rosenbaum and de Charms (1960), while studying the cathartic effects of vicarious aggression, found that the mere opportunity to respond to a frustrater lowered subsequent derogation of him, even where the original response was not an aggressive one (though this effect was present for low self-esteem subjects only).

The issue of catharsis and the cathartic effect is further confused by conflicting experimental findings. Aggressive behavior has been found to lower subsequent aggression in some instances (Feshbach, 1955; Thibaut, 1950; Thibaut & Coules, 1952) but in others has produced an increase in aggression (de Charms & Wilkins, 1963; Kenny, 1953). Feshbach himself (1956) obtained such findings in a later experiment and presented an interesting analysis of the aggression-enhancing or reducing effects of (aggressive) fantasy behavior. He hypothesized that aggression is cathartic where the subsequent situation is related to the situation involving the aggressive act, but that a second opportunity to aggress should lead to enhanced aggression where the two situations are unrelated.

Stagner (1944) argued that where aggression is the outcome of habitual hostility no cathartic effect is to be expected; presumably, since greater habit strength develops with repeated aggression, the strength of the latter should increase on those grounds. According to Buss (1961, p. 80 ff.), catharsis should decrease subsequent aggression where anger is present but increase it in the absence of anger because in the latter case inhibition has been lowered.

De Charms and Wilkins (1963) presented strong evidence that hostile statements engender subsequent heightened aggression and at the same time clarified an issue raised by Worchel (1960) dealing with the instigation-reducing effects of status restoration. In Worchel's experiment, the attacker and the status restorer were one and the same person, which made it impossible to separate the effects of status restoration from those of attacker's changed image as a supportive status restorer. From the Rosenbaum and de Charms (1960) results it might be inferred that status equalization, that is, reduction of the attacker's status, may serve as a substitute for victim's status restoration (see also Feshbach, 1964). De Charms and Wilkins (1963) found that status restoration by the subject himself or by a vicar was positively related to subsequent aggression, rather than negatively, as had been found by Worchel (1960). Again, one is tempted to infer that the social connotations of retaliation against a (deviant) attacker outweighed emotional aspects of accumulated anger or anxiety.

The evidence is thus truly conflicting. Aggression enhances or reduces subsequent aggression; a vicar's aggression may serve either as a substitute (Rosenbaum & de Charms, 1960), or as a model (de Charms & Wilkins, 1963) for the subject's subsequent aggression. However, the difficulties inherent in the interpretation of catharsis phenomena do not end here. After all, the cathartic effect may be due to any of the following reasons: (*a*) the subject has "let off steam" (an undifferentiated autonomic arousal model), (*b*) the subject has aggressed against someone or other (a hydraulic aggression model), (*c*) the subject has successfully aggressed against an indeterminate target (an expressive model), (*d*) the subject has successfully aggressed against the instigator (a self-esteem restoration or threat removal model), (*e*) the subject has seen the frustrater punished, even by someone else, and cognitive expectations regarding justice have been confirmed (a dissonance reduction model), (*f*) the subject has accumulated reactive inhibition or fatigue (a behavioral oscillation model), (*g*) the subject has developed guilt or aggression anxiety subsequent to aggression (an aggression-anxiety-with-delay model), (*h*) the subject perceives that a given amount of aggression is appropriate under the circumstances (a social reality model).

Where a cathartic effect is ostensibly absent, it may be the case that his aggressive habits have become stronger, that he perceives greater social desirability or tolerance for aggression, or that he perceives aggression to be the most effective way to deal with the situation, etc.

A series of experiments by Hokanson (1961) and his collaborators allows for interesting inferences regarding the role of expectations in determining the need for catharsis in order to produce the cathartic effect. Hokanson and Burgess (1962), for example, found that aggression against a high-status attacker did not reduce sympathetic arousal, while aggression against a low-status attacker did. These results appear to be directly relevant to the above mentioned variable of perceived or expected effectiveness of the attack.

Hokanson and Shetler (1963) found that when the subject is attacked by a high-status individual and does not expect an opportunity for retaliation his systolic blood pressure subsides at the same rate as when he was able to aggress. Though the authors discussed these results mainly in terms of substitute responses to frustration, such as withdrawal, an interpretation along the lines of perceived and—in the case of the high-status frustrater—accepted social expectation appears quite plausible.

In any event, Berkowitz' (1962, p. 196 ff.) exception to the absence of clear-cut interpretations to current experiments was well taken; too many variables have remained uncontrolled for any verdict regarding the notion of catharsis to remain undisputed or even to be meaningful.

To sum up, the study of the form of social interaction termed aggression has been hampered by two major conceptual difficulties:

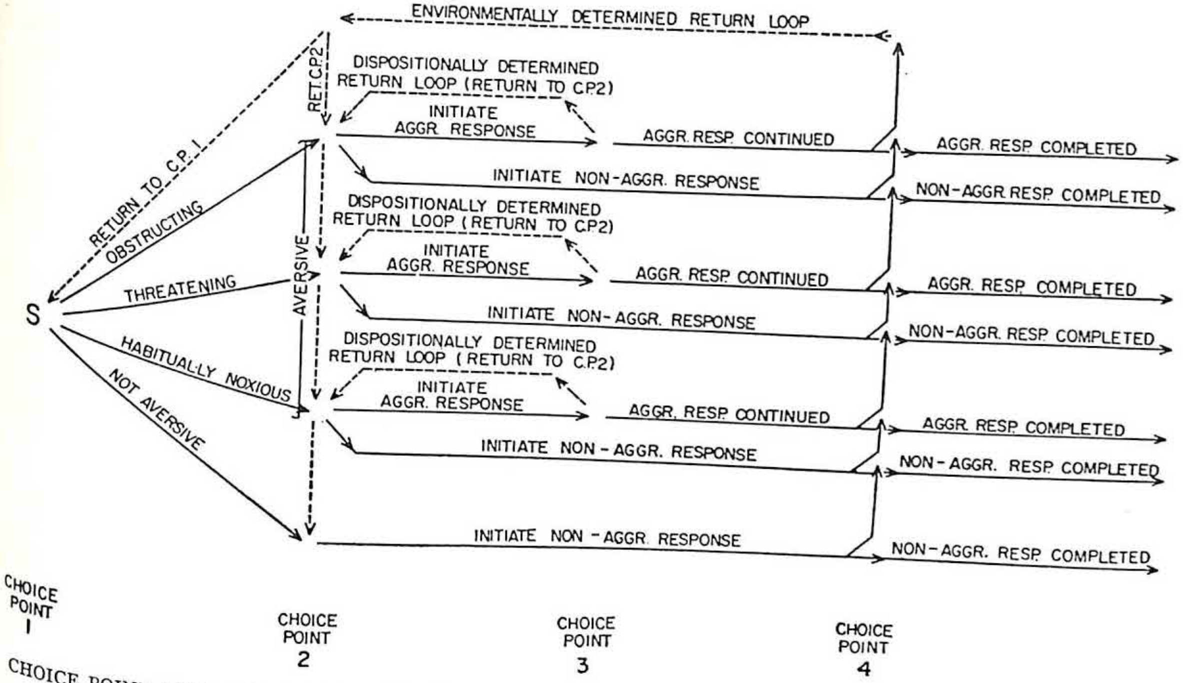
1. The standard terms originating in psychoanalytic theory and introduced to experimental psychology by the F-A hypothesis have been subjected to widely diverging definitions, some of which have resulted in attempts at defining what are obviously hypothetical constructs solely in behavioral terms.
2. The classification of superficially similar behaviors as one phenotype, disregarding or laboriously explaining different genotypes, which, were it not for a priori categories, might have proved far more amenable to experimental analysis.

One need not object to the use of such terms as aggression, displacement, catharsis, etc., as long as it is clear that these terms imply more than directly observable events. The admission of hypothetical constructs in the study of such complex social interactions as aggressive behavior need in no way detract from the rigor of empirical inquiry. In order to pursue such a study systematically, determinants other than those inherent in the stimulus itself must be investigated. While present assessments of habitual hostility as a personality variable may not allow for unequivocal predictions of aggression, there can be no doubt that personality variables do play a part in determining aggression, and interactions between the stimulus and such dispositional variables are probably going to afford even greater predictive accuracy than personality variables alone. The situational determinants facilitating or inhibiting aggression and those affecting the number of alternative nonaggressive responses are equally obvious in their relevance. Finally, an analysis which ignores the aggressor's expectations regarding the effectiveness of his projected act would probably ignore an important variable affecting the likelihood of any, not just an aggressive, behavioral act.

Keeping in mind the expected outcome of an aggressive act, we might then wish to investigate the dimensions along which a stimulus other than the original target may vary in order to present to the aggressor an acceptable alternative target. We may investigate whether the other traditional determinant of displacement, that of a lowered threshold, might be more economically explainable as a reduction in the number of perceived response alternatives. Catharsis and its effects can be studied in terms of the fulfillment of expectations regarding the completed act, and anticipation concerning subsequent aggression. Finally, as noted also by Feshbach (1964), there is a definite need for a precise analysis of the roles of emotional arousal in aggression, in terms of the various possibilities discussed earlier.

Summarizing Current Views on Aggressive Behavior as a Stimulus-Response Sequence

Feshbach (1964) presented a fairly simple flow chart for aggressive behavior. The Flow Chart for Aggressive Response suggests how this model can be amplified by considering the role of additional variables described earlier. Choice Point 1 denotes the classification of a stimulus as obstructing, threatening, habitually noxious—all defined as aversive or as not aversive. Where objective stimulus conditions are constant, the habit of classifying a member of certain stimulus subsets into the subjective subset of "noxious stimuli" may be viewed as being equivalent to the hostility syndrome, while the likelihood of perceiving a stimulus as threatening or obstructing might be seen as related to, respectively, self-esteem, and the ability to perceive alterna-



FLOW CHART FOR AGGRESSIVE RESPONSE

CHOICE POINT (CP) 1—Classification of stimulus
Dispositional determinants: optimism, self-esteem, hostility, transient states of perceptual orientation.
Autonomic state: Arousal may be conditioned to stimulus.
CHOICE POINT 2—Initiation of aggressive or nonaggressive response (fractional anticipatory goal response)
Dispositional determinants: aggressive tendency or aggressive habit.
Autonomic state: (*a*) Arousal may be classically conditioned to either aggressive or nonaggressive response and becomes established as an operant energizer. In the case of aggressive response, it may be labeled anger. (*b*) Arousal may be induced by cognitive mediation; the subject talks himself into being angry.
CHOICE POINT 3—Continuation of aggressive response, weakened by amount > 0, or return to CP 2
Dispositional determinants: conditional inhibition of aggressive responses (e.g., aggression anxiety).
CHOICE POINT 4—Continuation and completion of aggressive *or* nonaggressive response or return to CP 1 or CP 2
Situational determinants: social and other reinforcement cues.

tive goals or alternative goal paths to the original goal.

The role of autonomic arousal in general, not necessarily labeled as anger, at this point consists simply in eliminating alternatives of classification. More important than the autonomic state appears to be a person's momentary perceptual orientation at the time he is faced with the stimulus. He may be anticipating noxious, threatening, or obstructing events, and thereby increase the likelihood of classifying them accordingly. The effect of arousal at this point is, then, that of a probabilistic operator, which changes the likelihood that a stimulus will be classified in a certain way.

Choice Point 2 represents the initiation of a goal response. It is assumed that a stimulus categorized as not aversive has an initial probability of zero of evoking an aggressive fractional anticipating goal response, while noxious, obstructing, or threatening stimuli all have equal probabilities greater than (or equal to) zero of initiating an aggressive response. It is at this point that the construct "aggressive tendency" or "aggressive habit" becomes relevant. It plays the role of a probabilistic operator and, for reasons of economy, may be viewed as identical across classifications though, of course, differing across subjects.

Stimulus classification (at Choice Point 1), affected by both dispositional and situational determinants, can be conceptualized as standing in a multiplicative relationship to aggressive habit. Where a not aversive classification has been made *or* where aggressive habit is extremely low, the likelihood of an aggressive response being initiated becomes very small.

At Choice Point 1, anger may be present for two of the reasons discussed earlier. It may be an initially classically conditioned, later operant-arousal state associated with a response choice and subsequently labeled anger (see Shapiro, Crider, & Tursky, 1964), or it may be the result of deliberate self-stimulation, reinforced through previous, successful response activation; that is, the person talks himself into a rage (see Brehm, Back, & Bogdonoff, 1964). The role of autonomic arousal here is thus seen as simply energizing rather than as affecting the likelihood of a choice.

Choice Point 3 deals with the dispositional inhibitory variables such as aggression anxiety. Such aggression anxiety is seen as a conditioned inhibitory response which may either, in the traditional Hullian sense, detract from the reaction potential of the ongoing aggressive response, or, through a feedback loop, result in the selection of an alternative response. (The selection of an alternative response can also be accounted for in the traditional paradigm by referring to response hierarchies, where a new response becomes dominant if another one is inhibited. However, the notion of feedback and cognitive reevaluation of an ongoing response is a useful one in this instance.)

Choice Point 4 refers to the situational, including the social, determinants of the situation. The functions of these determinants may be viewed as permitting the person to mediate anticipated rewards or punishments which appear to be likely outcomes of his present ongoing response. The effect of these determinants is either to "clear" the ongoing response for completion, or to bring about a return to Choice Point 2 for selection of an alternate response. This process may take place where the original ongoing response is aggressive, that is, the person perceives the inappropriateness of his response and modifies it, or it may in some instances lead to the initiation of an aggressive response, even though the original response was nonaggressive. In the case where a stimulus had been classified as not aversive at Choice Point 1 and initiated a nonaggressive response, we should have the equivalent of attack against a target for purely social or similar reward expectations. It is important to note that repeated social aggression of this nature through classical conditioning may soon lead to a greater probability that a set of stimuli will be classified as noxious at Choice Point 1.

Choice Point 4 may also result in a return to Choice Point 1, that is, a reevaluation of the stimulus itself. Again, this reevaluation may take place by classifying a stimulus initially perceived as not aversive into one of the three aversive categories, or vice versa.

Once the aggressive or nonaggressive response has been completed, the subject's evaluation of the outcome may be assumed to take place. It is at this late stage that autonomic arousal or anger may occur or endure as a dissonance or anxiety reducer.

The present classification suggests that the dispositional determinants of perceiving stimuli in a certain manner (Choice Point 1) and strength of aggressive habit (Choice Point 2) should be susceptible to retraining, but such a retraining process could be quite lengthy. Building up inhibitions at Choice Point 3 may be a questionable strategy, since it implies less aggressive behavior, where aggressive response tendency is held constant, but also greater conflict. The opportunities represented by Choice Point 4 consist of making the likelihood of a favorable social or other payoff for aggression lower, or, preferably, of raising the expectation of a favorable payoff for a nonaggressive response. The general approach here has been either to induce a reevaluation of the stimulus, discussed by Pastore (1952), Pepitone (1958), and Feshbach (1964), or to encourage a new response to a stimulus which need not necessarily be perceived as less objectionable than before (Kaufmann & Feshbach 1963a, 1963b). Such a modification of an ongoing response may, but need not, be "frustrating," in that it is perceived as a threat or a path obstruction. As the second Kaufmann and Feshbach (1963b) study indicated, the determining factor may be whether a person refrains from aggression because he fears punishment or guilt, or because the rewards of nonaggression promise to exceed those of aggression.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.

Bandura, A., & Walters, R. H. *Social learning and personality development*. New York: Holt, Rinehart & Winston, 1963.

Beach, F. A. Bisexual mating behavior in the male rat: Effects of castration and hormone administration. *Physiological Zoology*, 1945, **18**, 390–402.

Beeman, E. A. The effect of male hormone on aggressive behavior in mice. *Physiological Zoology*, 1947, **20**, 373–405.

Berkowitz, L. The expression and reduction of hostility. *Psychological Bulletin*, 1958, **57**, 257–284.

Berkowitz, L. Anti-Semitism and the displacement of aggression. *Journal of Abnormal and Social Psychology*, 1959, **59**, 182–187.

Berkowitz, L. The judgmental process in personality functioning. *Psychological Review*, 1960, **67**, 130–142.

Berkowitz, L. Anti-Semitism, judgmental processes and displacement of hostility. *Journal of Abnormal and Social Psychology*, 1961, **62**, 210–215.

Berkowitz, L. *Aggression: A social psychological analysis*. New York: McGraw-Hill, 1962.

Berkowitz, L. Aggressive cues in aggressive behavior and hostility catharsis. *Psychological Review*, 1964, **71**, 104–122.

Bevan, W., Daves, W. F., & Levy, G. W. The relation of castration, androgen therapy, and pre-test fighting experience to competitive aggression in male C57BL/10 mice. *Animal Behavior*, 1960, **8**, 6–12.

Bindra, D. *Motivation: A systematic reinterpretation*. New York: Ronald Press, 1959.

Brehm, Mary L., Back, K. W., & Bogdonoff, M. D. A physiological effect of cognitive dissonance under stress and deprivation. *Journal of Abnormal and Social Psychology*, 1964, **69**, 303–310.

Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychological Bulletin*, 1951, **48**, 465–495.

Buss, A. H. *The psychology of aggression*. New York: Wiley, 1961.

Buss, A. H., & Durkee, A. An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology*, 1957, **21**, 343–349.

Buss, A. H., Durkee, A., & Baer, M. B. The measurement of hostility in clinical situations. *Journal of Abormal and Social Psychology*, 1956, **52**, 84–86.

Clark, G., & Birch, H. G. Hormonal modifications of social behavior. I. The effect of sex hormone administration on the social status of a male castrate chimpanzee. *Psychosomatic Medicine*, 1945, **7**, 321–329.

Cowen, E. L., Landes, J., Schaet, D. E. The effects of mild frustration on the expression of prejudiced attitudes. *Journal of Abnormal and Social Psychology*, 1959, **58**, 33–38.

de Charms, R., & Wilkins, E. J. Some effects of verbal expression of hostility. *Journal of Abnormal and Social Psychology*, 1963, **66**, 462–470.

Dollard, J., Doob, L., Miller, N., Mowrer, O., & Sears, R. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.

Estes, W. K. Toward a statistical theory of learning. *Psychological Review*, 1950, **57**, 94–107.

Feshbach, S. The drive-reducing function of fantasy behavior. *Journal of Abnormal and Social Psychology*, 1955, **50**, 3–11.

Feshbach, S. The catharsis hypothesis and some consequences of interaction with aggressive and neutral play objects. *Journal of Personality*, 1956, **24**, 449–462.

Feshbach, S. The function of aggression and the regulation of aggressive drive. *Psychological Review*, 1964, **71**, 257–272.

Gellerman, S. The effects of experimentally induced aggression and inhibition on word-association response sequences. Unpublished doctoral dissertation, University of Pennsylvania, 1956.

Guthrie, E. R. *The psychology of learning*. (Rev. ed.) New York: Harper, 1952.

Hokanson, J. E. Vascular and psychogalvanic effects of experimentally aroused anger. *Journal of Abnormal and Social Psychology*, 1961, **62**, 346–351.

Hokanson, J. E., & Burgess, M. The effects of status, type of frustration and aggression on vascular processes. *Journal of Abnormal and Social Psychology*, 1962, **65**, 232–237.

Hokanson, J. E., & Shetler, S. The effect of overt aggression on physiological arousal level. *Journal of Abnormal and Social Psychology*, 1961, **63**, 446–448.

Höss, R. *Kommandant in Auschwitz*. Autobiographische Aufzeichnungen des Rudolf Höss. Herausgegeben von M. Broszat. Munchen: Deutscher Taschenbuch Verlag G. m.b.H. & Co. K.G., 1963.

Janis, I. L. *Air war and emotional stress: Psychological studies of bombing and civilian defense*. New York: McGraw-Hill, 1951.

Kardiner, A., & Spiegel, H. *War stress and emotional illness*. New York: Hoeber-Harper, 1947.

Kaufmann, H., & Feshbach, S. Displaced aggression and its modification through exposure to anti-aggressive communications. *Journal of Abnormal and Social Psychology*, 1963, **67**, 79–83. (a)

Kaufmann, H., & Feshbach, S. The influence of anti-aggressive communications upon the response to provocation. *Journal of Personality*, 1963, **31**, 428–444. (b)

Kaufmann, H., & Marcus, A. M. Aggression as a function of similarity between aggressor and victim. *Perceptual and Motor Skills*, 1965, in press.

Kenny, D. T. An experimental test of the catharsis hypothesis of aggression. Unpublished doctoral dissertation, University of Washington, 1952.

Lesser, G. S. Extrapunitiveness and ethnic attitude. *Journal of Abnormal and Social Psychology*, 1958, **56**, 281.

Lindzey, G. An experimental investigation of the scapegoat theory of prejudice. *Journal of Abnormal and Social Psychology,* 1950, **45**, 296–309.

Liu, In-Mao. A theory of classical conditioning. *Psychological Review,* 1964, **71**, 408–411.

Maslow, A. H. Deprivation, threat, and frustration. *Psychological Review,* 1941, **48**, 364–366.

Maslow, A. H. Conflict, frustration, and the theory of threat. *Journal of Abnormal and Social Psychology,* 1943, **38**, 81–86.

McDougall, W. *An introduction to social psychology.* (Rev. ed.) Boston: Luce, 1926.

McNeil, E. B. Psychology and aggression. *Conflict Resolution,* 1959, **3**, 195–293.

Menninger, K. *Love against hate.* New York: Harcourt, Brace, & World, 1942.

Miller, G. A., Galanter, E., & Pribram, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

Miller, N. E. The frustration-aggression hypothesis. *Psychological Review,* 1941, **48**, 337–342.

Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *Journal of Abnormal and Social Psychology,* 1948, **43**, 155–178.

Miller, N. E., & Bugelski, B. R. Minor studies of aggression: II. The influence of frustrations by the in-group on attitudes expressed toward outgroups. *Journal of Psychology,* 1948, **25**, 437–442.

Pastore, N. The role of arbitrariness in the frustration-aggression hypothesis. *Journal of Abnormal and Social Psychology,* 1952, **47**, 728–731.

Pepitone, A. L. Attributions of causality, social attitude, and cognitive matching processes. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford Univer. Press, 1958.

Pepitone, A. L. *Attraction and hostility.* New York: Atherton Press, 1964.

Rosenbaum, M. E., & de Charms, R. Direct and vicarious reduction of hostility. *Journal of Abnormal and Social Psychology,* 1960, **60**, 105–111.

Rosenzweig, S. An outline of frustration theory. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944. Pp. 379–388.

Schachter, S., & Singer, J. E. Cogntive, social, and physiological determinants of emotional state. *Psychological Review,* 1962, **69**, 379–399.

Scott, J. P. *Aggression.* Chicago: Univer. Chicago Press, 1958.

Scott, J. P., & Fredericson, E. The causes of fighting in mice and rats. *Physiological Zoology,* 1951, **24**, 273–309.

Sears, R. R., Maccoby, Eleanor E., & Levin, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

Seward, J. P. Aggressive behavior in the rat: I. General characteristics, age and sex differences. *Journal of Comparative Psychology,* 1945, **38**, 175–197. (a)

Seward, J. P. Aggressive behavior in the rat: III. The role of frustration. *Journal of Comparative Psychology,* 1945, **38**, 225–238. (b)

Seward, J. P. Aggressive behavior in the rat: IV. Submission as determined by conditioning, extinction, and disuse. *Journal of Comparative Psychology,* 1946, **39**, 51–76.

Shapiro, D., Crider, A. B., & Tursky, B. Differentiation of an autonomic response through operant reinforcement. *Psychonomic Science,* 1964, **1**, 147–148.

Stagner, R. Studies of aggressive social attitudes: I. Measurement and interrelation of selected attitudes. *Journal of Social Psychology,* 1944, **20**, 109–120.

Stagner, R., & Congdon, C. S. Another failure to demonstrate displacement of aggression. *Journal of Abnormal and Social Psychology,* 1955, **51**, 695–696.

Steiner, I. D., & Johnson, H. H. Authoritarianism and "tolerance of trait inconsistency." *Journal of Abnormal and Social Psychology,* 1963, **67**, 388–391.

Thibaut, J. W. An experimental study of the cohesiveness of under-privileged groups. *Human Relations,* 1950, **3**, 251–278.

Thibaut, J. W., & Coules, J. The role of communication in the reduction of interpersonal hostility. *Journal of Abnormal and Social Psychology,* 1952, **47**, 770–777.

Tinbergen, N. Fighting and threat in animals. *New Biology,* 1953, **14**, 9–24.

Weatherley, D. Anti-Semitism and the expression of fantasy aggression. *Journal of Abnormal and Social Psychology,* 1961, **62**, 454–457.

White, R. K., & Lippitt, R. *Autocracy and democracy: An experimental inquiry.* New York: Harper, 1960.

Worchel, P. Hostility: Theory and experimental investigation. In, *Decisions, values and groups.* Vol. 1. New York: Pergamon, 1960. Pp. 254–266.

Zawadski, B. Limitations of the scapegoat theory of prejudice. *Journal of Abnormal and Social Psychology,* 1948, **43**, 127–141.

(Received January 25, 1965)

Psychological Bulletin
1965, Vol. 64, No. 5, 365–376

MEASUREMENT OF STIMULUS CONTROL DURING DISCRIMINATIVE OPERANT CONDITIONING[1]

HERBERT M. JENKINS
McMaster University, Ontario

The procedures in common use in discriminative operant conditioning are examined. It is concluded that they do not provide useful measures of stimulus control during acquisition because they confound the discriminative stimuli with antecedent reinforcement-nonreinforcement. Procedures which avoid confounding and provide measures of stimulus control during an uninterrupted acquisition process are described. 1 such procedure is applied in an experiment on sequential effects in discrimination learning. Substantial effects of antecedent reinforcement-nonreinforcement on generalized response probability and on the difference in probability of response to the discriminative stimuli are shown. The existence of effects of this type indicates that confounding can introduce serious biases into the measures of stimulus control in discriminative operant conditioning.

It is widely recognized that discriminative conditioning by the method of successive presentations occupies a central position in the analysis of operant behavior. It is therefore surprising that this form of discrimination learning has not been subjected to a more intensive experimental attack within the operant conditioning framework. In recent years, the technique of generalization testing developed by Guttman and Kalish (1956) has been used to gather a considerable harvest of new information on stimulus control as an end product of various conditioning regimes. Our knowledge of the *development* of stimulus control during conditioning, on the other hand, has not gone forward at a comparable pace.

The long-standing neglect of discrimination learning in operant conditioning may be due to an incompatibility between the procedures of operant conditioning, as they have evolved, and the conditions which must be met if stimulus control is to be measured during learning. Stimulus control is measured in terms of a difference in behavior attributable to experimental stimuli. Responding in the presence of a stimulus is not, in itself, evidence of stimulus control. Control is a synonym for covariation. For example, a generalization gradient shows stimulus control to the extent that some measure of the response varies with stimulus values (cf. Jenkins, 1965). In order to obtain an interpretable measure of stimulus control, one not confounded with other effects, it is necessary to measure behavior under conditions which differ only with respect to the stimuli in question.

Now, the thesis of the present paper is, in outline, as follows. The conditions in common use in operant studies do not make it possible to obtain the required measures of stimulus control during an uninterrupted process of acquisition. This is so because the procedures allow the effects of antecedent reinforcement and nonreinforcement on response probability to be confounded with the effects of the discriminative stimuli on response probability. The tactics which have been adopted to minimize or control for confounding have had bad side effects. These tactics lead to procedures which are, to put it moderately, awkward for analytical purposes. By modifying the procedures in certain ways, confounding can be avoided entirely; not just minimized or controlled for.

This argument is supported by an experiment in which some of the proposed modifications in procedure are applied. The specific purpose of the experiment is to trace the effect of antecedent reinforcement and nonreinforcement on stimulus control during

---

[1] This investigation was supported in part by Public Health Service Research Grant MH 08442-01 from the National Institute of Mental Health and in part by Grant APT-103 from the National Research Council of Canada.

learning. The results lend empirical support to the arguments about confounding since they show that effects of antecedent reinforcement and nonreinforcement, which would confound the measurement of stimulus control in typical operant procedures, do indeed exist in considerable strength.

HOW CONFOUNDING ARISES

Confounding arises whenever presentations of the discriminative stimuli are correlated with the context provided by prior reinforcement-nonreinforcement. Consider a hypothetical example in which the correlation is made obvious. Suppose one arranged 3-minute presentations of S_1 and S_2 and reinforced every response in S_1, and no responses in S_2. In effect, behavior in S_1 would be sampled repeatedly during the 3-minute S_1 presentation, and, with the exception of the interval between the onset of the stimulus and the first response, the samples would all be in the context of antecedent reinforcement. Responses in the S_2 period, on the other hand, would be sampled repeatedly and almost exclusively in the context of nonreinforcement. Therefore, context and stimulus presentation would be highly correlated.

Antecedent reinforcement-nonreinforcement can affect response probability on subsequent S_1 and S_2 presentations in several ways. For one thing, there are generalized effects of reinforcement and nonreinforcement. Reinforcement in S_1 increases the probability of response both to S_1 and to a related stimulus, S_2. Similarly, nonreinforcement in S_2 decreases response probability not only to S_2, but also to S_1. Thus, response probability both in S_1 and in S_2 would be expected to rise and fall in step with the sequence of positive (S_1) and negative (S_2) presentations. When presentations of S_1 and S_2 are correlated with the context of prior reinforcement-nonreinforcement, these generalized effects become inextricably mixed with the effects of the discriminative stimuli on responding.

A correlation between antecedent reinforcement-nonreinforcement and stimulus presentations introduces the possibility of another kind of effect. Reinforcement-nonreinforcement may develop a cue function. The basis for a cue function is evident in our hypothetical example. One reinforcement in S_1 predicts another reinforcement to follow, just as one nonreinforcement in S_2 predicts another nonreinforcement to follow. It is important to note, however, that confounding can occur even though no cue function develops. If there is some effect of context on generalized response probability and if context and stimulus presentation are correlated, confounding is present. The cue effect made possible by the correlation is just another way in which context can alter response probability and confound the measurement of stimulus control.

The use of regular reinforcement in our example makes the problem obvious, and no one would consider the performance differential between S_1 and S_2 periods to be a measure of stimulus control in such an arrangement. However, the problem remains, although it becomes less obvious, when the reinforcement to S_1 is intermittent, as it generally is in discriminative operant conditioning.

Why are operant conditioning procedures with maintained stimulus presentations more likely to introduce a correlation between context and stimulus presentation than are the procedures used with the runway, the jumping stand, and related devices? The reason is that in these other devices one does not continue to take measures or samples of behavior after the point at which the occurrence of reinforcement and nonreinforcement differentiates positive from negative trials. Successive samples are taken on separate trials. Since it is both simple and customary to arrange a random sequence of trials, the confounding of stimuli with antecedent reinforcement-nonreinforcement is quite naturally avoided.

What are the consequences of confounding? When confounding goes unrecognized and uncontrolled, differential responsiveness cannot be unequivocally attributed to the experimental stimuli. When it is recognized, tactics must be used to get around it. One tactic is to assess the role of the stimuli in extinction, thus eliminating reinforcement as a source of confounding. However, the procedure is obviously not suited to the continous measurement of the development of stimulus control because it interrupts conditioning by extinction in order to obtain a measurement. Another tactic is to use a control group in which only reinforcement differentiates positive from negative presentations (Frick, 1948; Smith & Hoy, 1954). However, this results in an indirect measure of stimulus control which rests on a between-group comparison of performance with and without the stimuli rather than on a response differential within an animal. The objection seems less serious if one finds that the reinforcement schedule alone produces no significant performance differential between positive and negative periods, but even in this fortunate circumstance the possibility that the reinforcement schedule introduces a bias when it interacts with the discriminative stimuli cannot be ruled out. Finally, a tactic so widely used that it may be considered a characteristic of discriminative operant conditioning is the use of intermittent rather than regular reinforcement. This tactic no doubt succeeds in reducing the magnitiude of confounding since it reduces the correlation between antecedent reinforcement-nonreinforcement and stimulus presentations. But, it does this at the cost of also reducing the correlation between the discriminative stimuli and reinforcement. The latter correlation, which is perfect under regular reinforcement, is an important and interesting variable in itself. A procedure in which the full range of the variable cannot be explored without giving up the possibility of a continuous measure of stimulus control during acquisition is seriously limited.

Ways to Avoid Confounding

In order to see how confounding may be avoided, it is necessary to examine in more detail the relation between context and discriminative stimuli both within stimulus presentations, and in the sequence of presentations.

Confounding within Presentations

Two ways of relating reinforcement to positive and negative stimulus presentations are shown in Figure 1. In Case A, S_1 and S_2 are of equal duration. On some or all S_1 trials, one or more reinforcements occur according to some schedule. The interval from trial onset to the first reinforcement (vertical arrow in Fig. 1) is labeled as the comparison interval, since any S_1–S_2 performance differential measured within this interval can only be attributed to the stimuli. Beyond this interval there is confounding, since performance must be regarded as a joint function of stimuli and antecedent reinforcement. Confounding from this source is very common in practice (e.g., Bower & Grusec, 1964; Dinsmoor, 1950; Frick, 1948; Pierrel & Sherman, 1962; Smith & Hoy, 1954). Commonly used schedules of reinforcement in S_1 also introduce the likelihood that reinforcement will serve a cue function. Reinforcement early in the period signals an increased probability of reinforcement in the remainder of the period. Consider, for example, a case in which S_1 and S_2 presentations are of equal duration and appear equally often. On S_1 trials, zero, one, or two reinforcements are programed with equal frequency; no reinforcements occur in S_2. For ease of calculation, assume the period of presentation to be divided into halves which contain, at most, one reinforcement. Now, ignoring the stimulus differences, the probability of reinforcement in the second half given none in the first half will be 1 : 9. However, when a reinforcement has occurred in the

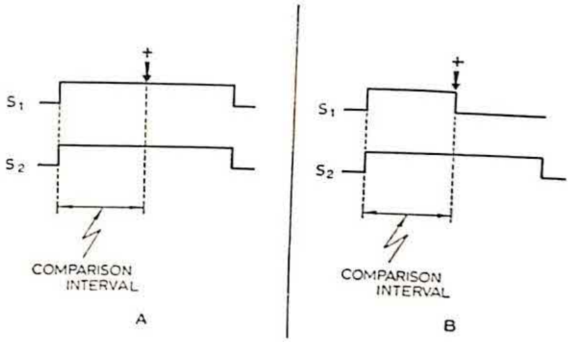


FIG. 1. Two ways of programing reinforcement (shown by vertical arrow) in relation to stimulus presentations. (In A, S_1 and S_2 trials are of fixed duration. In B, the S_1 trial terminates at the first reinforcement while the S_2 trial is of fixed duration. Within the comparison interval, performance is not confounded by reinforcement or nonreinforcement.)

first half, the probability of reinforcement in the second half jumps to 2:3. Thus, there exists a basis for the development of a discriminative or cue effect of reinforcement.

In Case B of Figure 1, the S_1 trial is terminated by reinforcement, but the S_2 trial is of fixed duration and extends beyond the comparison interval. The arrangement is no less confounded in principle than is Case A. A period of nonreinforcement exceeding a certain duration (when reinforcement is on an interval schedule) or continuing after the completion of a certain number of responses (when reinforcement is on a ratio schedule), occurs only on S_2 trials. The early experiments on discriminative conditioning by Skinner (1938, pp. 183–192) were on this plan. Another arrangement which would lead to S_2 durations in excess of S_1 is one in which the termination of the S_2 period is made contingent upon a period of no responding.

Confounding within trials may be avoided simply by terminating S_1 and S_2 according to the same rule or contingency. When this is done, all measurements will be taken within a comparison interval; that is, before the presentations are differentiated by the reinforcement contingency. For example, if S_1 is terminated in reinforcement by the completion of a certain number of responses (ratio contingency), then S_2 must also be terminated at the completion of this response requirement. Variable contingencies of the interval type can be used as well. Further, the probability that S_1 terminates in reinforcement when the contingency is met can be varied. Since the contingency will not always be met, particularly on S_2 trials, it is necessary that both S_1 and S_2 trials have some maximum allowable duration after which they are terminated by external control.

An arrangement which is ideal for the purposes of measurement may not be ideal for training a discrimination quickly. In particular, in order to avoid some form of secondary reinforcement from the termination of the trial, or just to allow for heavier doses of nonreinforcement, one might wish to maintain S_2 beyond the point at which the reinforcement contingency is met. Fortunately, it is a simple matter to avoid confounding even though trials are not actually terminated after the comparison interval. The requirement is only that the measurements taken in the comparison interval be separated from those taken beyond that interval. The control circuit which would have terminated trials after the comparison interval can be used to transfer the recording of data from one register to another, or simply to stop recording so that results for the comparison interval remain uncontaminated.

Confounding in the Sequence

It is clear that a procedure in which positive and negative periods of fixed durations are alternated (e.g., Bower & Grusec, 1964; Dinsmoor, 1950; Frick, 1948) confounds discriminative stimuli with the events of the immediately preceding presentation. Presentations of the positive stimulus, S_1, will be regularly preceded by nonreinforcement in the negative stimulus, S_2. Conversely, only S_2 presentations will be immediately preceded by reinforced responding in S_1. In a sequence of this kind, a difference in performance in S_1 and S_2 could arise from the carry-over effects of antecedent reinforcement-nonreinforcement in the prior presentations. A cue function based on reinforcement or nonreinforcement on the prior trial can also develop.

In order not to confound the measurement of stimulus control, the context given by preceding presentations must be the same for presentations of S_1 and S_2. The required independence of context is readily achieved when a random sequence of discrete S_1 and S_2 trials is used. The requirement can also be met in a continuous procedure, that is, one in which no third condition, such as removal from the apparatus, separates S_1 and S_2. One simply arranges a random sequence of presentations and programs the sequence without gaps. In a sequence of this sort, a stimulus remains unchanged for intervals which vary in duration in the same way as the lengths of runs of positive and negative trials vary in a random sequence of discrete trials. Accordingly, the frequency distribution of durations would be geometric. Variable stimulus durations were used by Pierrel and Sherman

(1962), but the distribution of durations was not geometric.

Even though a continuous procedure can be arranged to give contextual independence, it has some undesirable properties which can be avoided in discrete trial procedures. In the continuous procedure, any change in stimulus is accompanied by a change from a reinforced to a nonreinforced period, or vice versa. Therefore, the change of stimulus, rather than or in addition to distinctive features of maintained stimuli, may underlie a performance differential. Further, responses in the negative period receive secondary reinforcement at the transition to the positive period. The secondary reinforcement would be expected to slow the rate of reduction in S_2 responding, and perhaps also to raise the asymptotic level of S_2 responding. Breaking the stimulus sequence into separate periods avoids these difficulties and has the further analytical advantage of allowing experimental control over the temporal spacing of presentations. For these reasons, the discussion which follows is primarily concerned with discrete trial procedures.

There are several ways to separate stimulus presentations. The manipulandum may be retracted. A blackout may be used when the *S*s are pigeons, since darkness naturally induces inactivity in birds. An alternative is to separate trials by a discriminative stimulus clearly different from the trial stimuli (Jenkins, 1961, 1962). In the present experiment, for example, the pigeon's compartment remained illuminated at all times, but the response key was lighted only on the trials. Intertrial responses to the unlighted key, which occur at the outset of training, were quickly eliminated through nonreinforcement. The separation of trials in this way avoids the radical changes in the light-adaptation level of the pigeon's eye which blackouts cause. The method also ensures that responses are made to the trial stimulus, and not simply because a physical constraint or a stimulus having strong unconditioned effects has been removed. Finally, it may be noted that the use of a discriminative stimulus to separate trials makes operant and classical conditioning procedures more alike. The salivary response in classical conditioning, for example, is not physically blocked between trials. Rather, a discrimination between background alone, and background plus CS, keeps between-trial salivary flow at a minimum. An arrangement of this kind may be referred to as a discriminated trial procedure to distinguish it from what might be called a gated trial procedure in which the response is physically prevented between trials.

When discrete trials are employed, independence of context is easily ensured by the use of an entirely random sequence of S_1 and S_2 trials. A sequence of this type may, however, be inappropriate for the purposes of a particular experiment. One might require, for example, that the lengths of runs of trials of one type can be kept to a minimum. The Gellermann series (Gellermann, 1933) have a maximum run length of three trials of the same type and are sometimes used to determine the sequence of positive and negative periods in successive discriminations (e.g., Mabry, 1965; Pierrel & Sherman, 1962). These series were originally devised to avoid correct choices on the basis of position or alternation habits in simultaneous discrimination problems. Although they may be suitable for that purpose, their use in successive discriminations is unfortunate. The potential for confounding is evident, since only S_1 trials will appear after three consecutive nonreinforcements and only S_2 trials after three consecutive reinforcements. It should be noted that the animal does not have to "learn the series" in order for its performance to be dependent on the events of prior trials. Discriminative effects of reinforcement and nonreinforcement may of course develop, but even if they do not, the generalized effect of prior reinforcement or nonreinforcement on response probability would be a source of confounding.

It is of interest that sequences with special features may be contrived which, although highly nonrandom overall, yield unbiased measures of stimulus control at certain points. The sequence shown in Figure 2 is an example. The trial sequence generated by this schema, like the Gellerman series, has a maximum run length of three trials of the same type. There is, in fact, a regular alternation

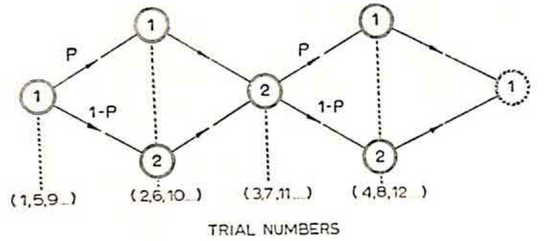


FIG. 2. Schema of a trial sequence in which odd-numbered trials alternate regularly between S_1 and S_2, while even-numbered trials provide unbiased measures of stimulus control. (S_1 is presented with probability P; S_2 with probability 1-P. The last trial, shown dotted, is the same as the first.)

between positive and negative stimuli on odd-numbered trials, and the sequence, taken as a whole, is confounded. The context given by the immediately preceding trial is equivalent for S_1 and S_2, but the context is not equivalent when two preceding trials are considered. An S_2 trial always occurs two trials in back of the odd-numbered S_1 trial, and, conversely, an S_1 trial always occurs two trials in back of the odd-numbered S_2 trial. On the other hand, on even-numbered trials, where the branches occur, S_1 and S_2 presentations are independent of context. Thus, by separating performance measures taken on even-numbered trials from the rest, unconfounded measures are obtained. The S_1 and S_2 performance differential on even-numbered trials following S_1 (Trials 2, 6, 10, etc.) might be systematically different from the differential on even-numbered trials following S_2 (Trials 4, 8, 12, etc.). That would indicate an effect of context on stimulus control. Whether or not the context has an effect, however, data from all even-numbered trials may be pooled without confounding as long as the probabilities of S_1 and S_2 trials following S_1 trials are the same as the probabilities following S_2 trials. In that case, there is no contingency between the context provided by the fixed trials and the discriminative stimuli on the branches. The effect of different contexts therefore applies equally to S_1 and S_2. The rule is that data taken from locations in a sequence at which antecedents are systematically different may be pooled without confounding as long as there is no contingency between antecedents and stimulus presentations.

Discrimination training procedures in which errors cause a repetition of the trial on which the error occurred always produce nonrandom sequences of stimulus presentations. Confounding can be avoided in such cases by a strategy similar in principle to the one discussed above. Data on repeat trials are separated from the rest so that they do not distort the measured effects of the discriminative stimuli. In the experiment to be reported, a procedure of this type was used.

EXPERIMENT ON SEQUENTIAL EFFECTS

It has been emphasized that, unless special care is taken in discriminative operant conditioning, the context given by antecedent reinforcement-nonreinforcement in the sequence of presentations, as well as within a presentation, may differ systematically for S_1 and S_2. The present experiment focuses on sequential effects in a discriminated trial procedure. Are there demonstrable effects of the prior trial on the current trial? What is the direction and magnitude of such effects? How do they depend on the amount of training?

Verplanck (1942) found strong sequential effects in discrimination learning in a runway experiment. He used a completely regular trial sequence consisting of repeated cycles of three negative trials followed by a single positive trial. Throughout training, running time on the negative trial which followed the positive trial was shorter than on negative trials which followed another negative trial. Early in training an effect in the same direction was found in terms of starting time, but the direction of the effect subsequently reversed so that starting times later in training were longer following a reinforced trial than they were following a nonreinforced trial.

It is of interest to go beyond the demonstration that prior reinforcement affects responsiveness and to ask how the S_1–S_2 performance differential depends on antecedent context. To answer the question, one must compare performance on S_1 and S_2 trials in the context of antecedent reinforcement and also in the context of antecedent nonreinforcement. The present experiment applies some of the previously discussed methodological devices to that purpose.

Procedure

Subjects. The *Ss* were 6 male White Carneaux pigeons 5-6 years old, without prior experimental history. They were maintained at 80% of their free-feeding weight.

Apparatus. An automatic key-pecking apparatus (Ferster & Skinner, 1957) was used. The compartment was constantly illuminated by diffuse light from a source which extended across the top of the response panel. Equipment sounds were masked by white noise from a speaker within the compartment. Reinforcement was access for 3.5 sec. to a tray containing mixed grain.

The stimuli were line grids which were automatically positioned behind the transparent key and transilluminated. The discrimination was based on the orientation of the grid. A vertical grid served as the reinforced or S_1 stimulus. The nonreinforced stimulus, S_2, was a horizontal grid, or in another phase of the experiment, a grid rotated 15° from the vertical.

Trials. A discriminated trial procedure was used. Trials were initiated by lighting the key and were terminated (key-light off) by the first response or by external control if no response was made within 5 sec. A response to S_1 was immediately reinforced. During the intertrial periods the key was dark, but the compartment remained illuminated. The intervals between onsets of trials were equally often 15, 30, or 45 sec. in random sequence. An intertrial response prevented the appearance of a trial within the next 60 sec.

Preliminary Training. The *Ss* were trained to respond to S_1 by the method of successive approximation. After several reinforced responses had been made, intertrial intervals were introduced and the maximum trial duration was shortened to 5 sec. The development of a quick response to the lighted key and the virtual elimination of intertrial responses generally required one or two sessions. Beyond the initial sessions, every *S* received 6 sessions of training, each consisting of 40 S_1 trials, before being exposed to S_2.

Trial Sequence. A type of correction trial procedure was used. An equal number of S_1 and S_2 trials was programed by the use of punched tape. Following correct trials (response to S_1 or no response to S_2), the tape advanced so that the next trial was governed by the taped random sequence. On the other hand, following incorrect trials (no response to S_1 or response to S_2) the tape remained in position and the same stimulus appeared on the next trial. Trials presented after an advance of the tape are called random trials to distinguish them from repeat trials.

Plan of Analysis. Random trials can only be preceded by a correct S_1 trial (S_1R), which entails a reinforced response, or by a correct S_2 trial (S_2NoR), which entails no response and no reinforcement. Performance on random S_1 and S_2 trials following each type of correct trial and performance on repeat trials were separately analyzed so that the effect of these antecedents could be identified. The analysis keeps track of only the immediately preceding trial.

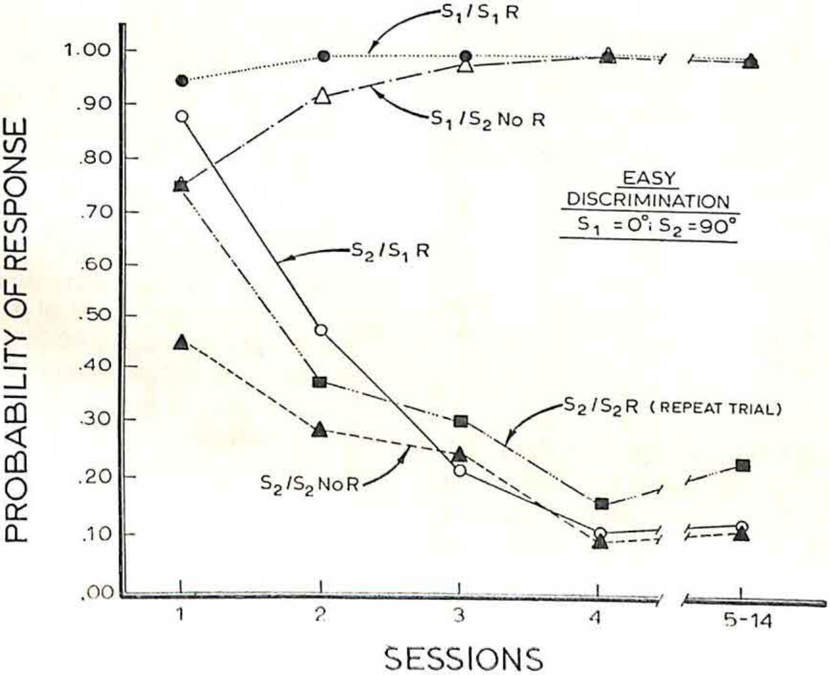


FIG. 3. Mean response probability by sessions as a function of stimulus and antecedent trial; easy discrimination. (Antecedent trial is shown after the slant line.)

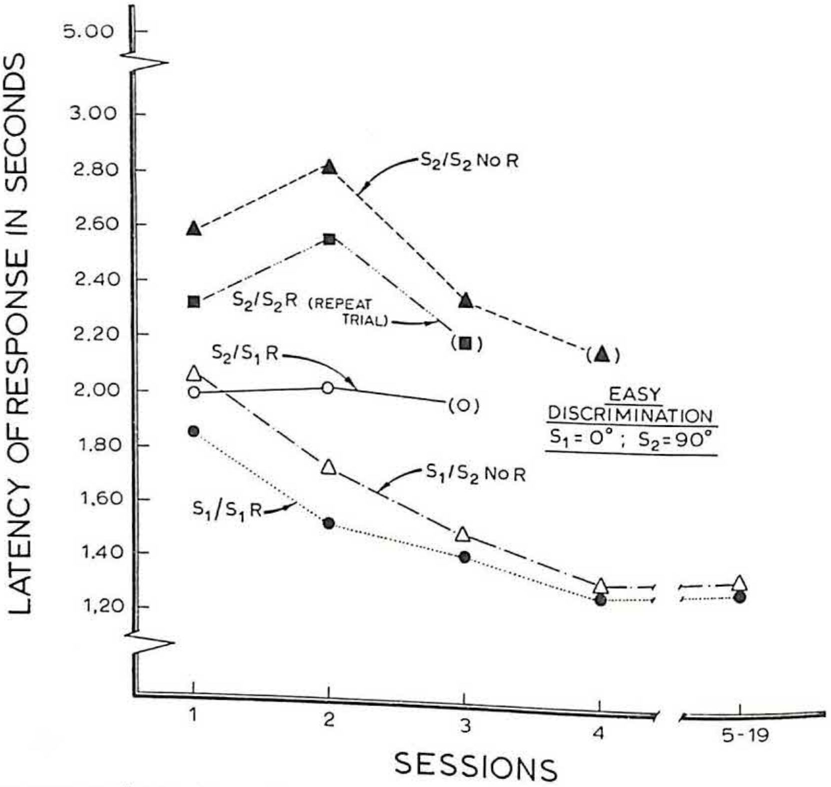


FIG. 4. Mean response latency by sessions as a function of stimulus and antecedent trial; easy discrimination. (Antecedent trial is shown after the slant line. Latencies become unavailable on S_2 trials as responses to S_2 diminish in frequency. Bracketed points are based on data from four or five of the six *Ss*.)

However, when there is substantial generalization from S_1, a correct S_2 trial will commonly occur only at the end of a series of repeat S_2 trials.

Two measures of performance were taken: the probability of response, and in the case of a response, its latency to the nearest tenth of a second.

Discrimination Problems. After preliminary training, 14 sessions were run on an easy discrimination. The S_1 was set at 0° (vertical) and S_2 at 90° (horizontal). In each session, 40 S_1 and 40 S_2 trials were programed. The *Ss* were then switched to a difficult discrimination: S_1 as before, S_2 at 15°. Seven sessions, each consisting of 20 S_1 and 20 S_2 programed trials, were then run.

Results

The probability of response to S_1 and S_2 following different antecedents is shown by sessions for the easy discrimination in Figure 3. Random S_1 and S_2 trials are preceded by S_1R (correct S_1 trial) or by S_2NoR (correct S_2 trial), giving rise to four separate curves. The fifth curve is for repeat S_2 trials; that is, an S_2 preceded by S_2R. Repeat S_1 trials were infrequent and are omitted. Corresponding latency data are given in Figure 4.

Early in training the antecedent trial appears to influence the response both to S_1 and to S_2. The mean probability of response was lower, and the mean latency of response was longer, following S_2NoR than following S_1R. Response strength on repeat S_2 trials was above the level for S_2 trials following an S_2NoR. The explanation of this difference probably lies in the fact that an S_2NoR typically occurs at the end of a run of repeat S_2 trials. The average number of repeat S_2 trials before an S_2NoR occurred was, for the first four sessions: 3.1, .8, .4, and .1. Thus, early in training, an S_2NoR occurs after a series of nonreinforced responses and is associated with a deeper state of extinction than is the average S_2R trial. The effects of antecedent trials diminish as the discrimination is perfected by continued training.

Mean latency and probability of response based on combined results for the first two sessions are given for random S_1 and S_2 trials as a function of antecedent in Table 1.

The antecedent trial significantly affected both the probability and latency of response to S_2. The difference in probability (.682 vs. .370) and the difference in latency (2.02 sec. vs. 2.71 sec.) were each significant at $p < .01$ by two-tailed t test. Although the effects of the antecedent trial on responsiveness to S_1 were in the same direction, they failed to reach significance either in terms of probability (.968 vs. .834, $p < .2$) or latency (1.69 sec. vs. 1.90 sec., $p < .2$).

Of central interest is the fact that the depressive effect of an S_2NoR was substantially greater on a subsequent S_2 than on a subsequent S_1. In terms of probability, the S_1–S_2 difference following S_1R was .968 − .682 or .286. Following S_2NoR it was .834 − .370 or .464. These differentials are significantly different from one another (.286 vs. .464, $p < .01$). Latency of response yields a parallel result. The S_1–S_2 difference in latency following S_1R was 1.69 sec. − 2.02 sec. or .33 sec. Following S_2NoR it was 1.90 sec. − 2.71 sec. or −.81 sec. Again, these differentials are significantly different (−.33 sec. vs. −.81 sec., $p < .02$).

TABLE 1

MEAN PROBABILITY OF RESPONSE, P, AND MEAN LATENCY, L, (IN SECONDS) ON RANDOM S_1 AND S_2 TRIALS AS A FUNCTION OF ANTECEDENT TRIAL

Antecedent trial	Stimulus	P	L
S_1R	S_1	.968	1.69
S_2NoR	S_1	.834	1.90
S_1R	S_2	.682	2.02
S_2NoR	S_2	.370	2.71

Note.—First two sessions, easy discrimination.

Response probabilities for the difficult discrimination are shown by sessions in Figure 5. In contrast with the easy discrimination, little or no improvement takes place. Also, unlike the case of the easy discrimination, the probability of response on repeat S_2 trials was indistinguishable from the probability of response to random S_2 trials following S_2NoR. However, the other major effects of antecedent trials, which were transient in the easy discrimination, were in this case maintained throughout training. Mean latency and probability of response for seven sessions are shown in Table 2. The direction of differences

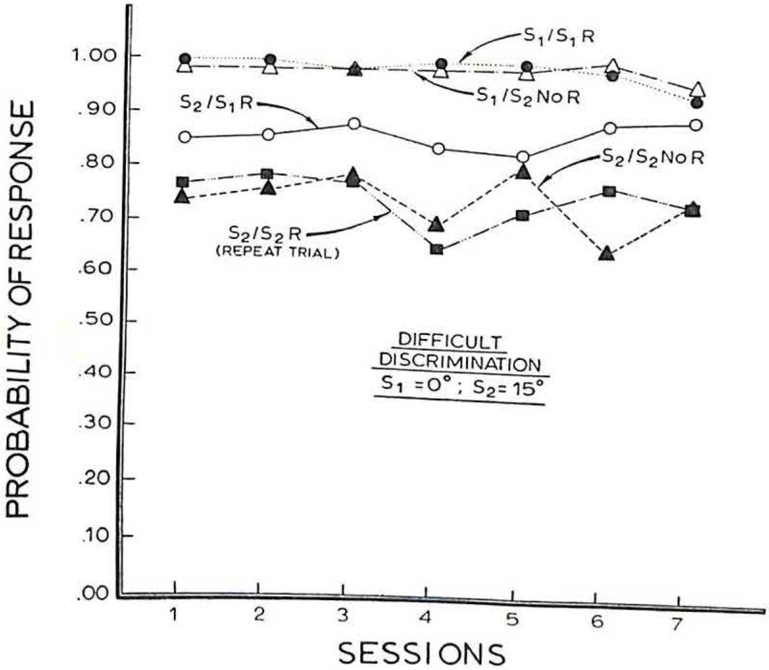


FIG. 5. Mean response probability by sessions as a function of stimulus and antecedent trial; difficult discrimination. (Antecedent trial is shown after the slant line.)

TABLE 2

MEAN PROBABILITY OF RESPONSE, P, AND MEAN LATENCY OF RESPONSE, L, (IN SECONDS) ON RANDOM S_1 AND S_2 TRIALS AS A FUNCTION OF ANTECEDENT TRIAL

Antecedent trial	Stimulus	P	L
S_1R	S_1	.987	1.44
S_2NoR	S_1	.978	1.59
S_1R	S_2	.866	1.66
S_2NoR	S_2	.744	1.92

Note.—Based on all sessions, difficult discrimination.

parallels the results for the easy discrimination. The antecedent trial had no significant effect on the probability of response to S_1 (.987 vs. .978, $t < 1$), but did show a significant effect on latency (1.44 sec. vs. 1.59 sec., $p < .05$). The antecedent trials had a clear effect on the probability of response to S_2 (.866 vs. .744, $p < .01$), and a detectable effect on latency (1.66 sec. vs. 1.92 sec., $p < .05$). In terms of probability, the S_1–S_2 difference for an antecedent S_1R was .987 −.866 or .121, while for an antecedent S_2NoR it was .978 − .744 or .234. These differentials are significantly different from one another (.121 vs. .234, $p < .01$). The corresponding differentials for latency differ in the same direction but not significantly (−.22 sec. vs. −.33 sec., $p < .2$).

In the early phase of the easy discrimination and throughout the difficult discrimination, the antecedent S_2NoR leads to a lower mean response probability and a longer mean response latency than does the antecedent S_1R. However, the depressive effect of S_2NoR on responsiveness to S_1 was relatively small and often insignificant, while the depressive effect on S_2 was more substantial. As a result, the discriminative stimuli exerted a greater degree of control over performance following S_2NoR than following S_1R.

Discussion

Antecedent reinforcement raised the probability of response to the stimuli presented next, and antecedent nonreinforcement lowered it. Parallel effects on latency were observed. There was no sign of effects in the opposite direction, although conceivably such effects could have arisen from frustration or by an inductive effect of the Pavlovian type (Pavlov, 1960, pp. 188–203). The results of Verplanck (1942) were, with one exception, in accord with the present results. The exception, it will be remembered, was that in his experiment the direction of the effect of antecedent reinforcement on one measure of performance (starting time) reversed during training. At the end of training, starting times were *longer* following a reinforced trial than following a nonreinforced trial. The reversal might have been due to the learning of a discrimination based in part on the regularity of the trial sequence which Verplanck used. However, the same trial sequence without the discriminative stimuli did not produce the reversal. Therefore, if regularity was indeed responsible, it must have been regularity in interaction with the differentiation of positive from negative trials by discriminative stimuli. In the present experiment, antecedent trials provided no cue to the nature of the next random trial, and the direction of the effect of the antecedent remained unchanged throughout extensive training.

These results have a clear bearing on method. Even in the complete absence of any cue properties in the trial sequence, substantial effects of antecedent trials appear. If presentations of the discriminative stimuli had been correlated with antecedents as they usually are in other procedures for discriminative operant conditioning, the resulting performance measures would have been a joint, confounded product of effects due to both antecedent events and the discriminative stimuli. Still larger effects would be expected from reinforcement and nonreinforcement within presentations, since within presentations there would be no intertrial interval to attenuate the effects. Antecedent trials affected not only the level of performance but also the S_1-S_2 performance differential. The differential was substantially greater following a negative trial than it was following a positive trial. How should this result be interpreted?

The antecedents were compound events, and the interpretation of how they exerted their effect must wait on experiments in which the parts of the compound are separated. One question such experiments should answer is: To what extent do the effects depend on the correlation of antecedent reinforcement and nonreinforcement with S_1 and S_2? It is quite possible that the effects are not entirely, or even largely, dependent on an associative or stimulus-specific process. They might have more to do with changes in the level of responsiveness as such. One might find, for example, that antecedent reinforcement or nonreinforcement on an irrevelant stimulus would have similar effects on the performance differential on the trial following. Perhaps just the opportunity to eat, even though not used as a reinforcement for a key peck, would raise responsiveness on the subsequent trial and reduce the performance differential. The existence of nonassociative effects of this kind is suggested by two findings: (*a*) relative stimulus-generalization gradients become steeper as drive is lowered (Jenkins, Pascal, & Walker, 1958), and (*b*) relative gradients also become steeper as the overall level of responsiveness decreases due to extinction in successive generalization tests (Friedman & Guttman, 1965; Jenkins & Harrison, 1960.

The present approach can be related in a general way to Harlow's error factor analysis of simultaneous-choice discrimination on the Wisconsin General Test Apparatus (Harlow, 1959). In both cases, the progress of discrimination learning is viewed as a composite of local effects. Further, it is perhaps more than coincidence that Harlow also finds better performance after a nonreinforced trial than after a reinforced trial. He identifies this error factor as a tendency to shift responses (response shift factor) from one object to another on successive trials. Thus an error is more likely to follow a correct choice than to follow an incorrect choice. His interpretation is specific to the simultaneous-choice task, but perhaps it will be found that a general lowering of responsiveness as a result of nonreinforcement plays a role both in successive- and in simultaneous-discrimination tasks.

SUMMARY

Procedures in common use in discriminative operant conditioning do not make it possible to obtain unbiased measures of stimulus control during an uninterrupted acquisition process. The difficulty is that in operant procedures, presentations of discriminative stimuli are almost invariably correlated with antecedent reinforcement-nonreinforcement. The correlation confounds whatever effects may be due to the discriminative stimuli with two different effects of antecedent reinforcement-nonreinforcement: (*a*) effects on the generalized strength of response and (*b*) cue effects.

Intermittent reinforcement is widely used in order to minimize confounding. However, intermittency of reinforcement is, in itself, a variable which undoubtedly affects the development of discrimination. A procedure which cannot provide interpretable measures of stimulus control throughout the range of this variable from partial to regular reinforcement is seriously limited.

One can avoid confounding within stimulus presentations by applying the same rule to the termination of positive and negative presentations. The rule or contingency which terminates positive trials and delivers reinforcement with some probability also terminates negative trials, but without reinforcement. Confounding can also be avoided even though the presentations are not actually terminated at the point at which the contingency is met. As long as performance measures which are taken up to that point are recorded separately from measures which are taken beyond that point, confounding will be avoided.

Confounding due to antecedents in the sequence of presentations can be avoided by the use of a random sequence of positive and negative stimuli. However, the sequence need not be random in its entirety. The requirement is only that at some points in the sequence the discriminative stimuli be presented at random and that the measurements taken at these points be separated from the measurements taken on the nonrandom portions of the sequence.

In principle, confounding can be avoided just as readily when successive presentations appear without a gap as when separate presentations or trials are used. However, separate presentations have certain analytical advantages. One way to separate trials is to train a discrimination between trial and intertrial (discriminated trial procedure) rather than to physically prevent intertrial responding (gated trial procedure).

An experiment in which the discriminated trial procedure was used showed clear effects of the antecedent trial on the probability and latency of response to the stimulus presented next. Responsiveness was increased by an antecedent positive trial and decreased by an antecedent negative trial. These effects were prominent early in training on an easy discrimination and were evident throughout training on a difficult discrimination. Sequential effects of this sort produce confounding when antecedent reinforcement-nonreinforce-

ment is correlated with the discriminative stimuli.

The antecedent trial affects not only the level of performance but the size of the differential in performance which is attributable to the discriminative stimuli. The differential is larger following a nonreinforced trial than it is following a reinforced trial.

The procedures which have been outlined for obtaining unbiased measures of stimulus control during acquisition are quite flexible. Presentations may be continuous or separated. Reinforcement may be regular or partial. Single or multiple responses may be required within each presentation. The contingency of reinforcement may be variable or fixed, and it may be based on time, response count, or any combination of these. Further, nonrandom sequences with any number of special constraints may be used as long as random presentations of the discriminative stimuli occur at some points in the sequence.

In view of the analytical advantages and flexibility of procedures which provide measures of stimulus control which are not confounded, one would hope that such procedures will become more popular in the study of discriminative operant conditioning.

REFERENCES

Bower, G., & Grusec, T. Effect of prior Pavlovian discrimination training upon learning an operant discrimination. *Journal of the Experimental Analysis of Behavior*, 1964, **7**, 401–404.

Dinsmoor, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *Journal of Experimental Psychology*, 1950, **40**, 458–472.

Ferster, C. B., & Skinner, B. F. *Schedules of reinforcement*. New York: Appleton-Century-Crofts, 1957.

Frick, F. C. An analysis of an operant discrimination. *Journal of Psychology*, 1948, **26**, 93–123.

Friedman, H., & Guttman, N. Further analysis of the various effects of discrimination training on stimulus generalization gradients. In D. Mostofsky (Ed.), *Stimulus generalization*. Stanford: Stanford Univer. Press, 1965. Pp. 255–267.

Gellermann, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *Journal of Genetic Psychology*, 1933, **42**, 206–208.

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 1956, **51**, 79–88.

Harlow, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959. Pp. 492–537.

Jenkins, H. M. The effect of discrimination training on extinction. *Journal of Experimental Psychology*, 1961, **61**, 111–121.

Jenkins, H. M. Resistance to extinction when partial reinforcement is followed by regular reinforcement. *Journal of Experimental Psychology*, 1962, **64**, 441–450.

Jenkins, H. M. Generalization gradients and the concept of inhibition. In D. Mostofsky (Ed.), *Stimulus generalization*. Stanford: Stanford Univer. Press, 1965. Pp. 55–61.

Jenkins, H. M., & Harrison, R. H. Effect of discrimination training on auditory generalization. *Journal of Experimental Psychology*, 1960, **59**, 246–253.

Jenkins, W. O., Pascal, G. R., & Walker, R. W. Deprivation and generalization. *Journal of Experimental Psychology*, 1958, **56**, 274–277.

Mabry, J. H. Discriminative functions based on a delay in the reinforcement relation. *Journal of the Experimental Analysis of Behavior*, 1965, **8**, 97–103.

Pavlov, I. P. *Conditioned reflexes*. (Trans. by G. V. Anrep) New York: Dover, 1960.

Pierrel, Rosemary, & Sherman, J. G. Generalization and discrimination as a function of the $S^{\Delta} - S^{\Delta}$ intensity difference. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 67–71.

Skinner, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.

Smith, M. H., & Hoy, W. J. Rate of response during operant discrimination. *Journal of Experimental Psychology*, 1954, **48**, 259–264.

Verplanck, W. S. The development of discrimination in a simple locomotor habit. *Journal of Experimental Psychology*, 1942, **31**, 441–464.

(Received May 6, 1965)

VOL. 64, NO. 6 DECEMBER 1965

Psychological Bulletin

DO COMPOUND EVOKED POTENTIALS REFLECT PSYCHOLOGICAL CODES? [1]

WILLIAM R. UTTAL

University of Michigan

The purpose of this paper is to examine the question of the utility of the compound evoked potential as a sign of perceptual experience. A discussion is presented of the origin of the peripheral nerve compound action potential and the brain potential evoked by impulsive stimuli. Some of the recent literature on the potentials from human Ss is considered to determine what progress has been made in correlating these electrical signs with psychophysical responses. It is concluded that while certain suggestions have been forthcoming, both denying and affirming the relation, the evidence is still inconclusive that simple correspondences exist between the 2.

Often in our attention to the details of experiment and communication, we tend to lose sight of the greater goals which motivate the particular research with which we are involved. For many of us in the field of physiological psychology, the main theme of our research has been a search for correlations between the functions of the brain and the verbal discriminative responses emitted by a subject in a psychophysiological experiment. So far, we have not progressed far towards this goal since not only do we lack adequate knowledge of both brain function and discriminative behavior, but there have been few experiments which have been specifically designed to compare the two types of data under identical stimulus conditions.

Nevertheless, many of us have assumed that these relations exist and that it is conceptually possible and practically probable by means of the currently available research methodology to establish them. It is, in addition, true that many of us have gone beyond this general act of faith to a more specific one. This more specific idea involves the assumption that it is possible to establish such relationships between the compound evoked potentials elicited by impulsive stimuli and psychological phenomena.

It is the purpose of this paper to explore this hypothesis, drawing upon the data forthcoming from the psychological and the physiological laboratories rather than from the philosophical exercises which originally gave rise to the mind-brain question.

One area in which there has been progress is in the comparison of electrophysiological and verbal responses to impulsive stimuli, such as clicks, flashes of light, and brief electrical shocks. Comparisons such as these are valid, since the stimulus conditions are well determined. Electrophysiological responses are time-locked with specific latencies following the stimulus presentation, and the psychophysical responses are controlled with an appropriate experimental design to reflect only the transient effects of the stimulus.

We will exclude from this discussion such other processes as intelligence, thought, and personality for which the stimulus conditions are less well understood. It is our opinion that these latter types of behavior will have to be contrasted with other physiological measures which, in general, are not yet available for laboratory investigation. However, it would be inappropriate to imply that these functions are more (or less) complex than responses to impulsive stimuli. A judgment of apparent magnitude of a simple stimulus is certainly a process of great complexity. The real issue is that there is, in fact, little physiological information and few proposed stimulus conditions which are relevant to the discussion of these "higher cognitive processes."

---

[1] The preparation of this paper and certain of the research projects described in it were supported in part by Research Grant MH 8786-01 from the National Institute of Health and Grant GB 2000 from the National Science Foundation.

In addition, we are also specifically excluding from this discussion nonevoked electrical potentials of the nervous system, such as electroencephalographic waves. Electroencephalographic waves are susceptible to uncontrolled variation, since we have not yet established all of the internal and external stimulus conditions which can modulate them. Vogel and Broverman (1964) reviewed the tenuous relationship which seems to hold between the electroencephalograph and the IQ, for example, without presenting a convincing argument that the electroencephalograph can be used as an indicator of intellectual competence.

The questions to which we will attend are well defined. It is important to establish exactly what we mean by a compound evoked potential (CEP) and to discuss current theories of its genesis. In particular, the relation between the CEP and the activity of single neurons must be elucidated. The second major purpose of this paper will be to present recent research from our laboratory and to review other relevant research which bears upon the question: Do measures of the compound evoked potential contain the same information that is expressed in the psychophysical judgments, and can we determine the nature of the code? An important limitation of this paper is that we shall be concerned mostly with electrophysiological data gathered from human subjects rather than from lower animals.

No paper on human CEPs would be complete without an acknowledgement of the contribution of G. D. Dawson. Dawson (1947) not only suggested and developed the averaging techniques for extracting the evoked cortical potential through the intact skull of a human being, but also developed the percutaneous procedure for recording the peripheral nerve compound action potential in man (Dawson & Scott, 1949). His work was extremely fruitful and can be credited with stimulating the widespread enthusiasm and effort in evoked potential work on the intact human being. The Dawson CEP techniques are particularly useful in comparative studies of psychophysical and neurophysiological responses carried out on human beings because they involve no surgery or anesthesia. It is this direct comparison between electrophysiological and psychophysical data usually collected from the same individual under identical, if not simultaneous, conditions with which we shall be most concerned. Indeed, because of Dawson's developments we today have at least the beginnings of a physiological description of psychophysical phenomena.

This paper supplements several of the reviews found in Koch (1962), particularly the article by Rosner (1962) which covers much of the relevant material up to that date. The reader's attention is also drawn to a recent major publication concerning cortical CEPs. The proceedings of the first major symposium on the sensory evoked response in man has recently appeared (Whipple, 1964) and contains much of the currently available knowledge pertinent to this paper.

Genesis of the Compound Evoked Potential

We shall, in general, be interested in two classes of CEPs in this part of the paper. The first is the action potential which is evoked from peripheral nerves by percutaneous stimulation and recording. An earlier paper (Uttal, 1960a) describes our modification of Dawson's techniques and the electronic circuitry we developed for the generation of these responses. Peripheral nerve action potentials are graded responses which vary in amplitude dependent upon a variety of stimulus intensity and timing considerations. These peripheral potentials are relatively strong signals and are far enough above the overall noise level so that they can be recorded and measured in single oscilloscope sweeps.

The second class of potentials with which we shall be concerned includes those potentials evoked presumably from the cerebral cortex and recorded through the intact skull and scalp. These potentials, unlike the peripheral nerve action potentials, are less synchronized, more attenuated, and, therefore, smaller in relation to the electrical and physiological noise in which they are imbedded. Thus, it is necessary to filter these signals by means of statistical techniques. The best known of these statistical filters is, of course, the averaging of time-locked signals such that events randomly distributed following the stimulus tend toward zero, while events which are related to the stimulus tend toward their average nonzero value. Both analog and digital devices have been applied as averagers.

Peripheral Nerve Action Potential

One of the first issues with which we must be concerned is the relationship between electrical responses to natural stimuli and the responses generated by impulsive electrical stimulation in peripheral nerves. Natural stimuli lead to relatively unsynchronized responses in the family of neural fibers which make up the compound nerve. Each response is a typical axonal all-or-none response and can be recorded with either microelectrodes inserted into an individual neuron or a gross electrode which records the activity of several cells on a single amplifier channel. Pulsed electrical stimuli, on the other hand, lead to a CEP which is made up of the responses of many nerve fibers all firing at the same instant. This latter type of response may be graded in amplitude, varying as more or fewer fibers contribute their signals to the response. The issue is: does the .5-millisecond-long electrical stimulus lead to a response which is coded in a different language and thus mislead the investigator about normal coding functions?

Gasser and Erlanger (1937) originally defined many of the properties of bundles of axons, such as the relation of nerve size to conduction velocity and temporal dispersion, on the basis of impulsive electrical stimulation and observation of the CEP. The question which is now raised is whether these evoked potentials reflect the same statistics of the nerve fiber population that could be obtained with natural stimuli.

In a very important recent paper, Casby, Siminoff, and Houseknecht (1963) reported the development of a cross-correlation technique which allowed them to repeat the earlier axonal studies but with natural stimuli. The exposed tibial nerve of a cat was laid over two electrodes, and a running cross correlation at various temporal delays was computed for the individual spike action potentials which were elicited by brushing, tickling, and/or pinching the skin. Although their ingenious technique itself is of considerable interest, for the purposes of this discussion the most important issue is that the results obtained from their study were essentially identical to those described by Erlanger and Gasser. This important finding means that the electrical pulse stimulus does not distort the code by the simultaneous firing of all of the axons. Rather, the synchronicity acts as a prestatistical processing which summarizes the coding scheme during, instead of following, the physiological process itself. Thus, the highly synchronized and graded compound action potentials recorded from the human peripheral nerves can be assumed to be an adequate indication of the normal coding function which would be observed if natural stimulus conditions were used.

It is also of considerable interest to consider whether or not repetitive stimuli of a relatively high frequency can give rise to low frequency oscillations in the responses of bundles of fibers. Such a demonstration would be interesting not only because of its relation to the peripheral nerve action potential, but also because it would suggest that we must consider that the summation of very brief spike action potentials from axons under certain conditions might also give rise to long-lasting waves even in central parts of the nervous system.

We have obtained just such results in our laboratory. The study about to be described was carried out with the equipment and method described in an earlier report (Uttal, 1960a) except that, instead of two or three pulses, the stimulus was a repetitive string of electrical pulses of equal interval and equal amplitude. We discovered that there were very long period oscillations in the amplitude of the compound action potential recorded

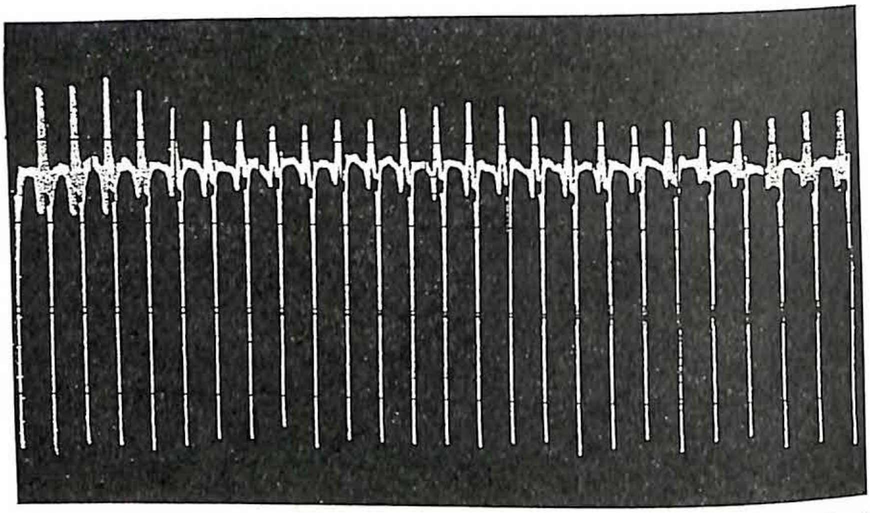


FIG. 1. A 200-millisecond-long sweep with stimuli (downward going spikes) every 8 milliseconds showing the oscillation of the amplitude of the compound potentials (upward going spikes) evoked from human peripheral nerves.

from human peripheral nerves when the nerves were stimulated with such a repetitive string. The effects of two independent variables, stimulus intensity and interstimulus interval, were investigated. A sample 200-millisecond record is shown in Figure 1. The downward going spikes are stimulus artifacts caused by pre-amplifier overdriving and serve only to indicate the time the stimulus occurs. The upward going lines are the CEPs from the peripheral nerves. A single CEP from a single stimulus is shown in Figure 2 on a 10-millisecond full-scale sweep. Thus, Figure 2 represents a time-expanded display of one of many sequential responses to the repetitive stimulation. The interpulse interval in the sample record of Figure 1 was 8 milliseconds. The stimulus was 3.5 milliamperes with a .5-millisecond duration. It can be seen that there is a long-term oscillation in the amplitude of the peripheral nerve action potential. The degree of this oscillation has been determined to be a function of the stimulus parameters mentioned above. Specifically, the oscillation occurs maximally (i.e., the highest peak-to-trough ratio) at lower intensities, even though the response amplitude increases with increasing stimulus intensity. The oscillation also appears to be affected by the interpulse interval of the stimuli, having a higher peak-

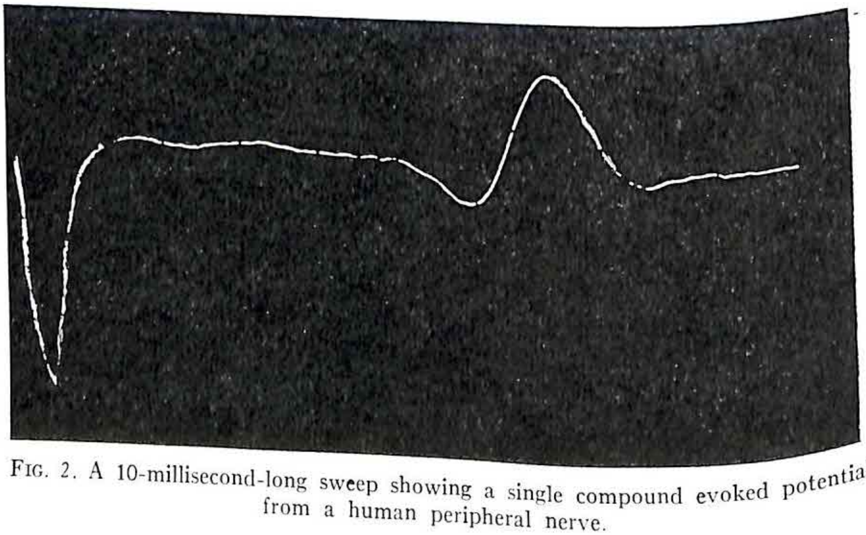


FIG. 2. A 10-millisecond-long sweep showing a single compound evoked potential from a human peripheral nerve.

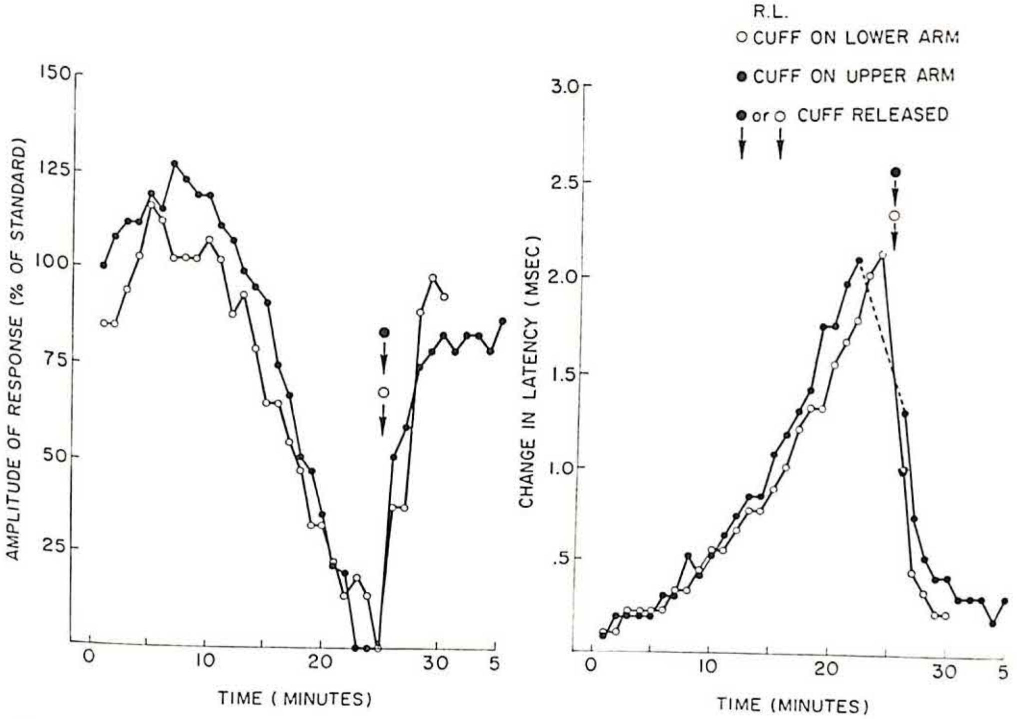


FIG. 3. The effect of ischemia produced by a pressure cuff on the amplitude and latency of the peripheral compound evoked potential as a function of time. (Note the small differences between the two conditions: cuff on lower arm below the recording electrode and cuff on the upper arm above the recording electrode.)

to-trough ratio when the pulses are close together, even though the wavelength of the oscillation does not appreciably change.

The most important fact for this discussion is that the response amplitude of a bundle of axons can under certain conditions be modulated to produce wave forms which are far longer lasting than the individual spike action potentials of the neurons in the bundles. It is also interesting to note that the wavelength of these responses appears to be of the order of 100 milliseconds, and that the wavelength appears to be independent of both interpulse interval and stimulus amplitude.

The origin of the oscillations in the amplitude of the peripheral nerve action potential has not yet been explained. Two principal hypotheses remain once we exclude artifacts of technique and accept these oscillations as true neurophysiological events. The first assumes that the oscillation is due to the changes in the nerve membrane potentials which occur during tetanic stimulation. Thus, the oscillation could reflect the collective "hunting" by the involved nerve membrane potentials for new equilibrium levels. The oscillation would be related in this case to refractory periods and a nonmonotonic accumulation of after potentials during the repetitive train of stimulations.

The second hypothesis assumes that the oscillation is not a function of the nerve membrane dynamics of the stimulated nerves, but rather is a result of reflexively generated centrifugal activity. Thus, in this case, efferent activity would be assumed to be interfering with the afferent activity represented by this compound action potential.

The existence of such reflexive activity has been established. Toennies (1938) has described a dorsal root reflex of sensory fibers in which spike action potentials are reflected back along the same afferent fibers which conveyed the nerve impulse away from the point of stimulation. Fullerton and Gilliatt (1965) have more recently reported another axon reflex in human peripheral nerves which appears to be produced at branch points in efferent fibers stimulated antidromically. The fact that the effect reported in this present experiment was more pronounced at the low stimulus intensities at which sensory fibers are more likely to be excited (Dawson, 1956) suggests that the dorsal root reflex is a more likely explanation, if a feedback explanation ultimately proves to be more appropriate.

Although the origin of the oscillations of successive responses is not clear, we can be relatively specific about the nature of the individual CEP evoked from human peripheral nerves. Changes in the amplitude of the peripheral CEP seem to be almost entirely a function of the number of single axon responses which are summed to make up the integrated response. This holds true under even the most extreme environmental conditions as illustrated by a recent study from our laboratory (Uttal, 1964) in which a sphygmomanometer cuff was used to create an ischemic condition in the arm. The study used the same paradigm described above—electrical pulse stimuli were used to stimulate major trunks of peripheral nerves, and the CEP was recorded from a superficial point of the nerve trunk further up the arm. The pressure block was imposed in one experiment above the recording electrode and in another experiment below the recording electrode. In each case, the reduction in nerve response amplitude at the recording point was the same as a function of time (see Fig. 3). In the case in which the cuff was below the recording electrode, the blood supply of the nerve was normal at the recording point.

Since the amplitude of response of a nerve fiber should be normal in a region of a nerve which has normal metabolic conditions, even if it has arrived through an abnormal region, this result suggests that the reduction in peripheral nerve CEP is due to the reduced number of individual neuronal responses. Thus the reduction in amplitude is due to heightened thresholds at the ischemic stimulating point rather than to reduced action potential amplitude at an ischemic recording point.

Because so much of the variance in CEP amplitude is accounted for by the number of responses, even in this extreme ischemic environment, it is not inappropriate to conclude that the amplitude of a peripheral nerve CEP in a given human subject is a very direct measure of the number of firing neurons. There is, of course, a great deal of literature on this subject from experiments with animals. The special problem here is that with the human subject we are dealing with responses generated and recorded percutaneously, and data such as these help us to understand the nature of this special type of recording situation.

Compound Evoked Brain Potential

From the preceding discussion it appears that we have a fairly good theory about the genesis of the peripheral nerve CEP. The situation is not so clear when we discuss the CEP which is evoked by impulsive stimulation of peripheral nerves and recorded through the intact skull of the human subject.

One of the major issues concerning brain CEPs is the origin of the long lasting and slowly changing recorded voltages. In 1959, Purpura comprehensively reviewed theories of the genesis of cortical potentials and stated that the early axonal models which were invoked to explain slow brain potentials could no longer be accepted, principally because spike activity could not be associated experimentally with the slow potentials. These models, he further stated, have been replaced by newer ideas in which slower dendritic and synaptic potentials are hypothesized as the basic subunits which are summated to produce the slow electrocortical potentials.

Amassian, Waller, and Macy (1964) have also expressed this same point of view in the recent symposium on sensory evoked response in man. For a number of reasons, including the experimental disassociation of spike and slow activity mentioned above and considerations of the volume recording technique, they also suggested that the spike action potentials do not contribute directly to the CEP recorded from the surface of the head. They also stated that the slow potentials already known to exist, such as the postsynaptic and after potentials, are added together to produce the wave form.

Much of their evidence supporting such a viewpoint is in terms of the analysis of a small sample of neuronal action potentials.

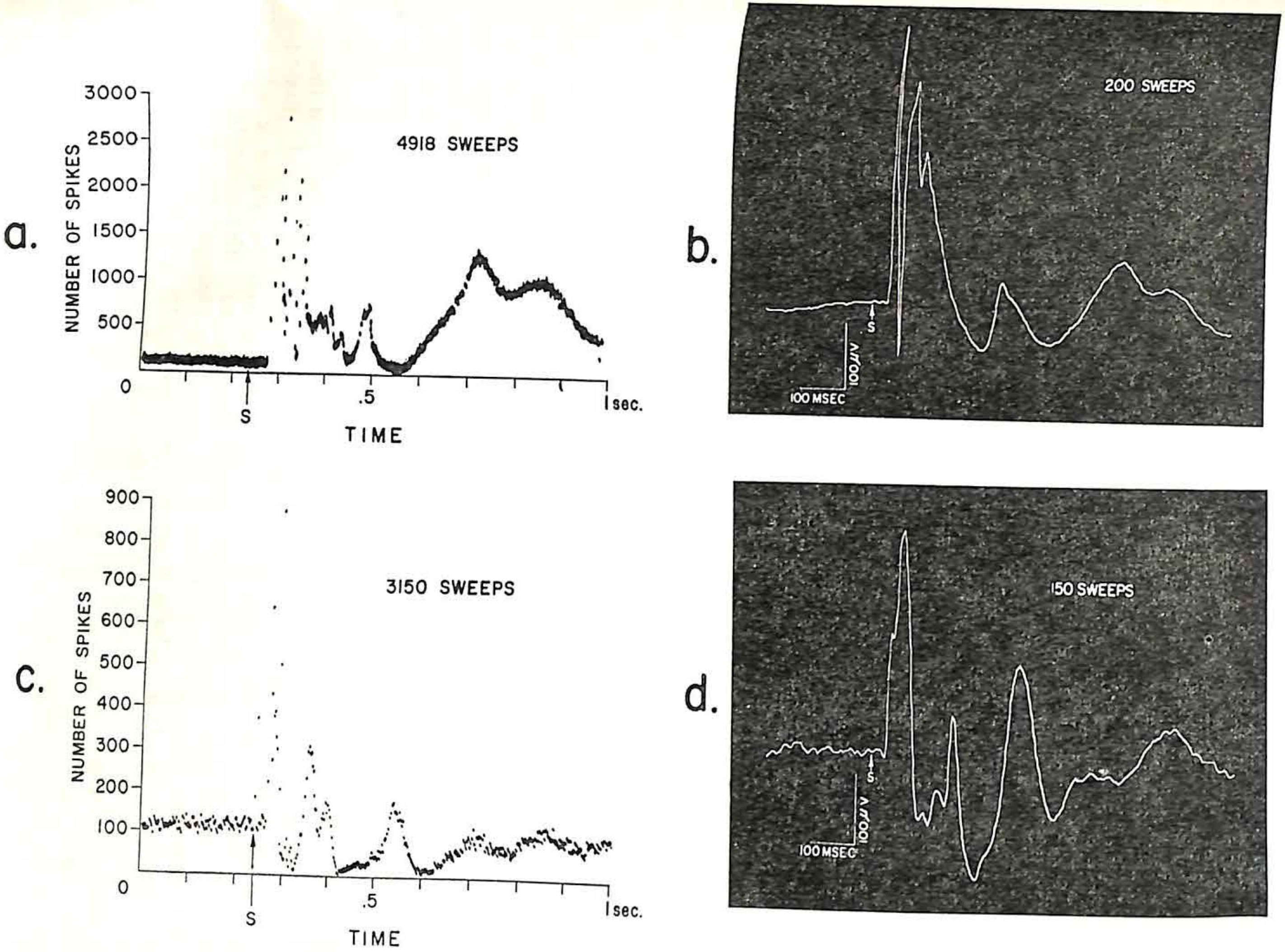


FIG. 4. This figure (reproduced with the author's permission, from Fox & O'Brien, 1965) shows the striking similarity between the compound potential evoked from the visual cortex of the cat and a latency histogram of the spike action potentials from a single cell. (Copyright 1965 by the American Association for the Advancement of Science.)

Modern computing equipment allows much larger sample sizes to be collected, and, on the basis of these large samples, relations do emerge between spike activity and the slow potentials.

Fox and O'Brien (1965) used a combination of computer analysis techniques to demonstrate a very high correlation between the frequency histograms of latencies of a single cortical cell as recorded with an extracellular microelectrode and the averaged CEP as recorded through the same electrode but with an amplifier passing lower frequency signals. The relationship between these two measures is rather astounding (see Fig. 4) and indicates that spike action potential may have been somewhat prematurely rejected as a major contributor to the slow potentials. Fox and O'Brien's work suggests that the CEP is a measure of the probability of spike firing and demonstrates the missing relation between spike activity and the CEP. The work on the amplitude oscillation of peripheral nerve CEP described above also demonstrated a link between the slow response and the spike action potential, regardless of the origin of the oscillation.

The axonal and the slow dendritic or synaptic explanations may, however, not be quite as separate as they seem. Synaptic and dendritic potentials may be themselves very closely related to the spike activity, being precursors of spikes with the same encoded information. We may, therefore, find ourselves replicating the history of atomic physics wherein two seemingly contradictory theories were vigorously debated only to be found, in later years, to be essentially different expressions of the same information. Thus, the CEP might become a useful sign of brain activity conveying a summated form of the information carried by both the spike and the graded activity of cortical neurons. From the point of view of the psychologist, the specific origin of the potentials then would become a question of lesser importance.

For the time being, let us consider only that the genesis of the cortical CEP is still being vigorously debated and that while some feel that it strongly relates to synaptic and dendritic graded activity, challenges have been inserted into the discussion which argue strongly for direct spike-action-potential contributions to the cortical CEP.

Another controversy concerning the genesis of the cortical CEP revolves around the question of whether these potentials are cortical at all. Bickford, Jacobson, and Cody (1964) have proposed that many of the components of the evoked potential (particularly parts of the responses to acoustic stimulation) are not neurogenic, but represent activity of muscles time-locked to the stimulus. Using a deceptively simple experiment, they suggested that major components of the evoked cortical potential are indeed not cortical but arise from the neck muscles. On the other hand, recent studies such as the one by Domino, Matsuoka, Waltz, and Cooper (1964) demonstrate that potentials recorded through the intact skull are indeed identical in shape to those recorded from the exposed surface of the brain, although greatly reduced in amplitude. On the basis of this type of evidence, we now feel that Bickford's warnings, though prudent advice at the time they were expressed, have been counterindicated by newer direct evidence.

Brazier (1964) has cautioned that the averaged evoked potential as a recording measure may be misleading, even though it is demonstrably cortical in origin. For example, the average of a very large amplitude signal which is varying in latency to an extent greater than the length of the response could lead to an averaged response with zero amplitude.

Another caution which has not often been expressed involves the variability of the cortical potential associated with day-to-day or session-to-session variations which affect serial measurement. The computation of an averaged evoked potential is a tedious and time-consuming process, and the general procedure has been to record the averaged response to, say, 80 presentations of the first stimulus, alter the stimulus, and then take the averaged response to another 80 stimuli. The recording and averaging for a single record could, therefore, take several minutes if sufficient time is given between stimuli to minimize sequential effects. Only a few measurements are then possible during each experimental run, and uncontrolled changes occurring during the experimental session may actually introduce a variability into the serial records greater than that produced by the stimulus variations. Counterbalanced order of stimulus presentation can, in part, solve this problem, but many investigators probably ignore this issue of sequential effects.

We have carried out experiments in which the effect of simply allowing a subject to sit in the experimental room and respond to a constant stimulus over a period of an hour was measured. Figure 5 is a series of photographs of records from a subject in which each photograph is the plot of 100 sequential averaged evoked responses. The decline in the amplitude of portions of the wave as a function of time is evident, even though the stimulus conditions (.5-millisecond pulse duration, 7-milliampere stimulus intensity) were constant. This result implies that it is necessary to have an acompanying response which is developed during the same time period as the responses to the test stimulus. This standard response has been obtained in our laboratory by averaging alternative records for two different stimulus conditions.

A standard stimulus, constant throughout the entire experiment, is presented on odd-numbered stimulus presentations while the test stimulus condition, varying from one step of an experiment to the next, is presented on even-numbered stimulus presentations. The computer controlling the experiment averages the even-numbered and the odd-numbered responses separately. A comparison is then made between the two averaged records to determine the effects of the stimulus.

The averaged evoked potentials are, therefore, collected during an identical time period for both the standard and the test stimulus. The price that one pays for this is to reduce even further the number of measurements that one can make during a single experimental session. On the positive side, though, is the fact that measurements can be made in terms of deviations from the standard response.

Such a technique should solve many problems, but the psychophysiologist still is faced with the problem of deciding which of the many parameters of the complex wave form of the CEP he should measure. Certainly latency of the various components is one significant measure, but so are other measures, such as amplitude and area under the curve relative to some arbitrary zero potential. We shall not defend any specific measure, since convenience, relevance to the particular study, and simple intuition all contribute to the evolution of a satisfactory measure. And, indeed, the determination of which particular measure one uses is the very essence of the inquiry we are pursuing in this paper. We are examining these potentials to determine if they do, by some measure or another, relate closely to the psychological code. To select one or the other at the present time would only beg the question.

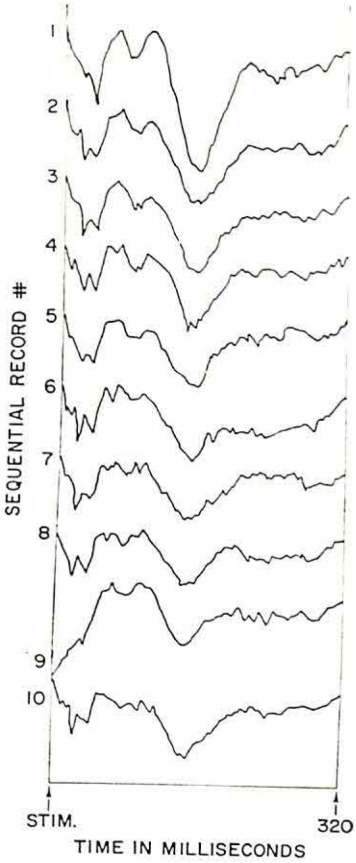


FIG. 5. Sequential averaged compound evoked cortical potential taken over a period of an hour with constant stimulus conditions to demonstrate the variability of the brain compound evoked potential without variation in the experimental stimuli.

EVOKED POTENTIALS AND PSYCHOLOGICAL CODING

Over the last few years, a number of investigators have explored the problems of the evoked potentials in search of functional relationships between stimuli and the variations of the CEP. We shall not be concerned with this physiological data, but in this short review we shall deal only with those experiments which directly compare some aspect of human thought or behavior with some apparently equivalent electrical potential. The study of the physiological parameters is, of course, basic to a full understanding of the latter class of experimental findings, and the reader is directed to the work of Dawson (1947) and Rosner and his colleagues (Allison, 1962; Goff, Rosner, & Allison, 1962), in particular, for the foundation necessary to a full understanding of the physiological organization of these responses.

In an earlier work (Uttal, 1960a) on the comparison of the peripheral nerve CEP and the psychophysical judgment of apparent magnitude, we concluded that the amplitude (i.e., count of the number of activated fibers) was the critical value in the encoding of subjective intensity. However, to the best of our knowledge no study on the relation of the peripheral CEP to the quality of sensation has yet appeared in the literature, and there is, therefore, nothing to say about this type of encoding.

A number of papers discuss the relation of the brain CEP to various aspects of psychological performance. The final sections of this report will review the evidence for a relation between the brain CEP and the stimulus and several psychological parameters.

The association between certain of the aspects of the brain CEP with the meaningfulness of the stimulus pattern has been investigated by a number of recent experimenters. Meaningfulness has several different definitions in this group of experiments.

In the first (Haider, Spong, & Lindsley, 1964), meaningfulness was defined in terms of the discrimination of a visual stimulus which was slightly less intense than the others which occurred before and after. The subject was instructed to respond to the randomly interspersed less intense stimuli by pressing a key. During the time period following the stimuli, signals from macroelectrodes on the head were recorded on an analog tape recorder for later analysis. The details of the experimental design are not germane to this discussion, but the general results were: (*a*) The amplitude of the evoked potential declined over the 1.5-hour period of the experiment; (*b*) There was a corresponding decline in the percentage of correctly detected signal stimuli; (*c*) On the basis of a separate averaging of those electrical responses to stimuli which were missed and those which were correctly identified, it was found that responses to correctly identified stimuli were associated with CEPs larger than the CEPs with which responses to incorrectly identified stimuli were associated.

The authors of the report conclude that the CEP may serve as a relatively satisfactory measure of the state of vigilance of the subject. This experiment is not convincing, however, for although there may indeed be a correlation between the amplitude of the evoked potential and the state of vigilance, it seems likely that the attenuation and enhancement of the CEP may be an epiphenomenon unassociated with the intellectual activity they seek to measure. We have shown (see Fig. 5), for example, that the attenuation in the amplitude of the evoked potential goes on under conditions in which the stimulus does not change at all. Furthermore, it would seem that Bickford's argument for myographic contamination of the CEP would be most applicable in just this situation, for attentiveness might well be associated with a generalized orientation reaction which could be more directly measured by myographic recordings from other parts of the body. Most damaging to their argument, however, is that pupil dilation was not measured or controlled. Pupil size, varying with accommodation or general alerting reaction, could influence the CEP in the direction indicated in their experiment without a direct relation to the cortical processes involved in vigilance. Such an experimental situation is also extremely susceptible to technical artifacts due to small shifts in the location or contact resistance of the recording electrodes with scalp and head movements.

Somewhat more convincing, though still subject to the same sort of criticism, is a similar study by Davis (1964). In this study the meaningfulness of an acoustic stimulus pattern was related to evoked potentials in a vigilance task. The stimuli were tones which might or might not vary in intensity. It was the subject's task to determine whether or not the third of four tone bursts was louder than the second. The first and fourth served as "ready" and control signals, respectively. The separation between each of the four tone bursts was 2.5 seconds. The results of the experiment indicated that the averaged CEP response to those stimuli in which a decision task was called for was larger than when no decision was called for. The subject had been instructed in certain of the trials not to make a discriminative judgment, but simply to push the response button each time he heard anything or, alternatively, not to attend at all to the experimental stimuli, but rather to read a magazine during the experiment.

A criticism of this experiment is, of course, that the time between stimuli was also long enough to allow the subject to orient in some secondary fashion, that is, to develop a new "attentive state." It may be that the measure of attentiveness or vigilance suggested by both of the two preceding reports is an indirect effect generated either by myographic contamination or by some indirect physiological attenuation of the signal. It would have been no more surprising if the CEP had actually been larger with lower levels of detection performance, just as the alpha rhythm of the electroencephalograph increases when intellectual or perceptual activity is reduced.

There is, however, a third study in which the association between meaningful and nonmeaningful stimuli and the amplitude of the CEPs is more rigorously demonstrated. Chapman (1964) described an experimental paradigm in which irrelevant influences were more carefully controlled. His subjects did mental arithmetic, adding two numbers which were presented on a single visual display preceded by a noninformative warning light and with another noninformative field separating the two numbers. All four of the visual stimuli occurred within a second as the displays changed from one to the next in a quarter of a second. This rapid action reduced greatly the amount of time for changes in muscular set or other orienting reactions. In this experiment with appropriately counterbalanced practice order, the evoked responses to informational stimuli are significantly greater than the evoked responses associated with noninformational stimuli.

This experiment adds credence to the argument of the two weaker studies on the effect of meaning or attentiveness on the CEP from the brain. In this case, more adequate control had been effected which strengthened the argument that the enhancement of the evoked potentials for significant stimuli is a true effect associated with the neural bases of the related intellectual processes, rather than an epiphenomenon produced by only incidentally related preparatory responses. Unfortunately, too little is known of the latencies of preparatory responses to make this experiment crucial.

Among the first investigators to make a direct comparison between psychophysical measures and CEPs were Geisler, Frishkopf, and Rosenblith (1958). They found that the evoked response produced by an auditory click stimulus appeared in the averaged CEP at approximately the same stimulus intensity as that at which the subject first reported he was able to detect the stimulus.

Rapin (1964) has also found a moderately high degree of relation between the audiometric threshold and the threshold for the averaged evoked potentials. These are important results because they, along with the phenomenon of saturation demonstrated in somesthesis (i.e., medium level stimuli produce maximal electrical potentials, Uttal & Cook, 1964), suggest a wide degree of spatial divergence of the induced excitation as it proceeds up the ascending pathways. The temporal course of the CEP, of course, is also very much prolonged compared with the duration of the stimulus, and this suggests that a great degree of temporal dispersion also exists. Nevertheless, the Rapin and Geisler et al. studies are illustrative of a very important correlation between the CEP and the psychophysical judgment at the threshold level.

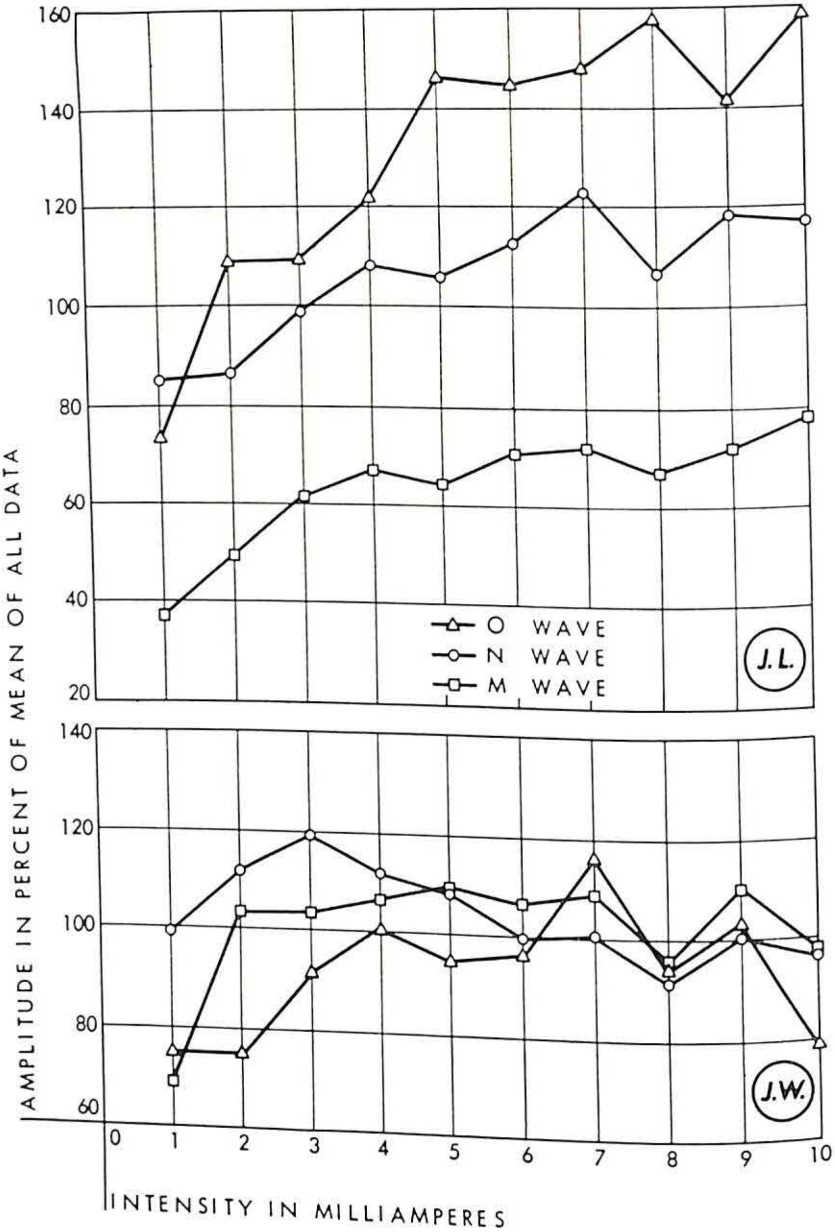


FIG. 6. The relation between stimulus intensity and amplitude of the brain compound evoked potential. (Note that the amplitude reaches near maximum values at low stimulus values, thus demonstrating the phenomenon of saturation of amplitude.)

The saturation phenomenon mentioned above does not, however, correspond to psychophysical judgments of higher stimulus levels. Figure 6 shows the amplitude of the evoked potential for a typical subject as a function of stimulus intensity, and Figure 7 shows the related monotonic increase in the subjective magnitude estimates (from Uttal & Cook, 1964). Since the CEP amplitude does not change while the subjective amplitude is varying widely, simple amplitude measures do not appear to be adequate measures of subjective intensity. If there had been any systematic change at all, we might have been close to defining some coding relation between the psychophysical and the electrophysiological variables.

This then brings us to another important

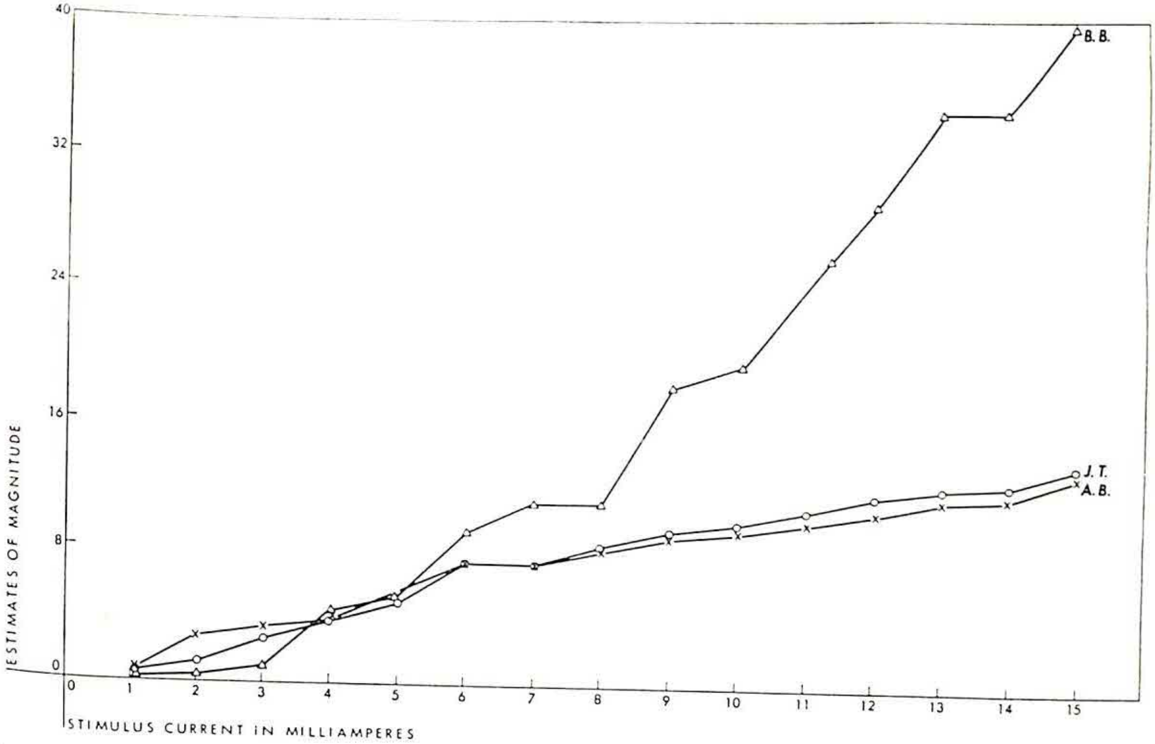


FIG. 7. The relation between psychophysical estimates of apparent magnitude and the stimulus intensity. (Note the "monotonic" increase of apparent sensory magnitude as stimulus intensity increases.)

point. Rosner (1962) has pointed out that the functional relationship between the psychophysical and electrophysiological variables need not be a linear one. Certainly the spatial representation of the body in the somatosensory area of the brain is better described as a topologically constant relation than as a linear one. In this scheme, the body image is stretched and distorted from the relative dimensions of the body as measured externally. This one example shows that there is no a priori reason to assume that the relationship between the neural message and the psychophysical code is a linear one and that any such assumption of linearity begs the question this field of research aims at answering.

In a different sort of study, investigators searched for a CEP correlate of the masking phenomena found in near simultaneous electrical stimulation of the finger tips. Matin (1959) has shown the temporal function for the masking, while our laboratory (Uttal, 1960b) has investigated the interaction as a function of distance (number of fingers) between the test and masking stimuli. One of the major results of the latter study was that there was no masking for stimuli applied to different hands, but that extensive masking occurred between fingers on the same hand. An examination (Uttal, 1964) of the brain CEP recorded under analogous situations has shown this same effect to obtain at least qualitatively in that no interaction could be demonstrated between the stimuli applied to opposite sides of the body. This finding is contrary to the results for the cat in which there was strong inhibitory interaction between stimuli applied to opposite paws (Nakahama, 1958).

Barlow (1964), in a difficult and complex experiment, studied the relation of the time at which visual stimuli were perceived to the components of the evoked potential. The stimulus in this experiment was the sudden movement of a spot on the face of a cathode ray tube. His findings were that the perception of the movement occurred within about 15 milliseconds of the movement, but that the first components of the CEP appeared much later, 80 to 90 milliseconds after the stimulus. Apparently, he did not detect any specific primary early evoked responses, such as those which are found in the cortical somatosensory CEP. The difficulties in experiments of this sort are formidable, and Barlow stressed that his conclusions were severely limited by the ambiguity in measuring the time of visual perception. Nevertheless, his conclusions indicate a lack of correlation between the CEP and the time of perception.

Halliday and Wakefield (1963) have studied evoked potentials to cutaneous electrical stimuli in patients with lesions in the ascending tract of the spinal cord or brain stem which resulted in a loss of position or cutaneous sensation. They reported that the cortical evoked potential was not affected in cases in which the cutaneous senses were absent, but was interfered with when the joint-position sense was absent. They tended, therefore, to attribute the evoked potential in its entirety to the joint-position sense and further concluded that it is carried by the posterior column pathways.

Donchin, Wicke, and Lindsley (1963) reported a study in which the visual brain CEP evoked by two sequential flashes, a weak stimulus followed after a variable length of time by a bright masking stimulus, seemed to reflect an analogous subjective response. In the analogous psychophysical experiment, it was found that a second stimulus blanked out the perception of an earlier one, if the earlier one occurred within a certain range of delays (0–25 milliseconds). The brain CEPs followed a pattern which the investigators felt was the same. They went on to conclude, therefore, that the CEPs appeared to be simple algebraic sums of the responses produced individually by the test and masking stimuli:

> Within limits, therefore, the data suggest that the electrocortical activity associated with perceptual interactions to paired flash stimulations may result from additive brain processes. This tends to support the assumptions of other workers who have used paired somesthetic stimuli [p. 1286].

However, this is exactly what those who have worked with paired somesthetic stimuli have not found. Allison (1962), Shagass and Schwartz (1962), and Uttal and Cook (1964) all found that the second of two evoked responses was diminished in size during the 100 milliseconds following the initial stimulus, and that these temporal inhibitions led to vast deviations from a simple additive process. The difference between this group of somesthetic experiments and the results of Donchin, Wicke, and Lindsley may lie in the fact that the stimuli used in the latter experiment were so vastly different in amplitude. The intensity of the test flash was .25 millilamberts while the intensity of the masking flash was 260 millilamberts. Even if one considers the saturation phenomenon, the high intensity of the blanking flash would tend to swamp out the effect of the first 100 milliseconds.

CONCLUSION

The inconclusive findings of the few experiments in the literature which made direct comparisons of psychological and electrophysiological responses in man restrict the answer to the question asked in the title of this paper. Certainly, there has to be some relation between nervous activity and behavior, but equally as certainly we have not yet established the specific measures we should make to "decode" the sensory experience. Perhaps there could not be a single crucial experiment that would answer the query for us completely, because different sensory mechanisms may be encoded in different languages.

This brief survey does indicate, though, that an increasing amount of effort is being expended in investigating this problem and that many investigators have made the initial assumption that there does appear to be a decodable relationship. In peripheral nerves, particularly (this means that we must consider the somatosensory system exclusively, since the other receptor nerve tracts lie deeply and for the present, inaccessibly buried in the head), there seems to be a high degree of association between amplitude response and subjective magnitude estimates. Quality coding, on the other hand, has not yet been tested, primarily because of the vagueness of defining the cutaneous modalities and the difficulties of working with natural stimuli.

The great unanswered issue lies in the relation of the CEP recorded from the sensory cortex to psychological experience. It would be presumptuous to imply that any general answer can yet be made to this question. The issue is whether or not this statistical function—the CEP—has any association with mental activity or is some irrelevant epiphenomenon. The data are inconclusive. Several studies suggest that there is an association, while others seem to show no correlation, when simple measures of amplitude and latency are tabulated. Unfortunately, as a group, the experiments which explore the relation of what we have called stimulus meaningfulness exhibit insufficient controls, weak stimulus situations, and ambiguous results.

Much additional research is needed. It is, however, important to remember that a new element has been inserted into this endeavor. It is now possible to carry out electrophysiological observations on the intact and normal human being while simultaneously making psychophysical inquiries. Although the suggestion of continued research may seem trite, Dawson's creative contributions now allow us to study both aspects of this question in its most appropriate experimental animal, man. For psychologists, in particular, this is a crucial development.

REFERENCES

Allison, T. Recovery functions of somatosensory evoked responses in man. *Electroencephalography and Clinical Neurophysiology,* 1962, **14**, 331–343.

Amassian, V. E., Waller, H. J., & Macy, J., Jr. Neural mechanism of the primary somatosensory evoked potential. *Annals of the New York Academy of Sciences,* 1964, **112**(Art. 1), 5–32.

Barlow, J. S. Some preliminary observations on evoked responses and perception of visual stimuli in man. *Electroencephalography and Clinical Neurophysiology,* 1964, **17**, 468–469.

Bickford, R. G., Jacobson, J. L., Cody, D., & Thane, R. Nature of average evoked potentials to sound and other stimuli in man. *Annals of the New York Academy of Sciences,* 1964, **112**(Art. 1), 204–223.

Brazier, Mary A. B. Evoked responses recorded from the depths of the human brain. *Annals of the New York Academy of Sciences,* 1964, **112**(Art. 1), 33–59.

Casby, J. U., Siminoff, R., & Houseknecht, T. R. An analogue cross-correlator to study naturally induced activity in intact nerve trunks. *Journal of Neurophysiology,* 1963, **26**, 432–448.

Chapman, R. M. Evoked responses to meaningful stimuli while solving problems. Paper read at Psychonomic Society, Niagara Falls, Canada, October 1964.

Davis, H. Enhancement of evoked cortical potentials in humans related to a task requiring a decision. *Science,* 1964, **145**, 182–183.

Dawson, G. D. Cerebral responses to electrical stimulation of peripheral nerve in man. *Journal of Neurology, Neurosurgery, and Psychiatry,* 1947, **10**, 134–140.

Dawson, G. D. The relative excitability and conduction velocity of sensory and motor nerve fibres in man. *Journal of Physiology,* 1956, **131**, 436–451.

Dawson, G. D., & Scott, J. W. The recording of nerve action potentials through skin in man. *Journal of Neurology, Neurosurgery, and Psychiatry,* 1949, **12**, 257–273.

Domino, E., Matsuoka, S., Waltz, J., & Cooper, I. Simultaneous recordings of scalp and epidural somatosensory-evoked responses in man. *Science,* 1964, **145**, 1199–1200.

Donchin, E., Wicke, J. D., & Lindsley, D. B. Cortical evoked potentials and perception of paired flashes. *Science,* 1963, **141**, 1285–1286.

Fox, S., & O'Brien, J. Duplication of evoked potential waveform by curve of probability of firing of a single cell. *Science,* 1965, **147**, 888–890.

Fullerton, Pamela M., & Gilliatt, R. W. Axon reflexes in human motor nerve fibres. *Journal of Neurology, Neurosurgery, and Psychiatry,* 1965, **28**, 1–11.

Gasser, H. S., & Erlanger, J. *Electrical signs of nervous activity.* Philadelphia: Univer. of Pennsylvania Press, 1937.

Geisler, C. D., Frishkopf, L. S., & Rosenblith, W. A. Extracranial responses to acoustic clicks in man. *Science,* 1958, **128**, 1210–1211.

Goff, W. R., Rosner, B. S., & Allison, T. Distribution of cerebral somatosensory evoked responses in normal man. *Electroencephalography and Clinical Neurophysiology,* 1962, **14**, 697–713.

Haider, M., Spong, P., & Lindsley, D. B. Attention, vigilance, and cortical evoked-potentials in humans. *Science,* 1964, **145**, 180–182.

Halliday, A. M., & Wakefield, G. S. Cerebral evoked potentials in patients with dissociated sensory loss. *Journal of Neurology, Neurosurgery, and Psychiatry,* 1963, **26**, 211–219.

Koch, S. (Ed.) *Psychology: A study of a science.* Vol. 4. *Biologically oriented fields: Their place in psychology and in biological science.* New York: McGraw-Hill, 1962.

Matin, E. *Temporal aspects of cutaneous interaction with two-point electrical stimulation.* Unpublished doctoral dissertation, Columbia University, 1959.

Nakahama, H. Contralateral and ipsilateral cortical responses from somatic afferent nerves. *Journal of Neurophysiology,* 1958, **21**, 611–632.

Purpura, D. P. Nature of electrocortical potentials and synaptic organizations in cerebral and cerebellar cortex. In C. C. Pfeiffer & J. R. Smythies (Eds.), *International review of neurobiology*. New York: Academic Press, 1959. Pp. 47–163.

Rapin, Isabelle. Evoked responses to clicks in a group of children with communication disorders. *Annals of the New York Academy of Sciences*, 1964, **112**(Art. 1), 182–203.

Rosner, B. S. Psychophysics and neurophysiology. In S. Koch (Ed.), *Psychology: A study of a science*. New York: McGraw-Hill, 1962. Pp. 280–333.

Shagass, C., & Schwartz, M. Cerebral cortical reactivity in psychotic depressions. *Archives of General Psychiatry*, 1962, **6**, 235–242.

Toennies, J. F. Reflex discharge from the spinal cord over the dorsal roots. *Journal of Neurophysiology*, 1938, **1**, 378–390.

Uttal, W. R. The three-stimulus problem: A further comparison of neural and psychophysical responses in the somesthetic system. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 42–46. (a)

Uttal, W. R. Inhibitory interaction of responses to electrical stimuli in the fingers. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 47–51. (b)

Uttal, W. R. *The effect of ischemia on the peripheral nerve action potential in man and its relation to somatosensory magnitude coding.* (Mental Health Institute reprint 125) Ann Arbor: Univer. of Michigan, 1964.

Uttal, W. R., & Cook, Louella. Systematics of the evoked somatosensory cortical potential: A psychophysical-electrophysiological comparison. *Annals of the New York Academy of Sciences*, 1964, **112**(Art. 1), 60–80.

Vogel, W., & Broverman, D. M. Relationship between EEG and test intelligence: A critical review. *Psychological Bulletin*, 1964, **62**, 132–144.

Whipple, H. E. (Ed.) Sensory evoked response in man. (Art. 1) *Annals of the New York Academy of Sciences*, 1964, **112**.

(Received April 22, 1965)

Psychological Bulletin
1965, Vol. 64, No. 6, 393–401

CONVERGING OPERATIONS FOR PERCEPTUAL DEFENSE

THOMAS NATSOULAS[1]
University of California, Davis

A discussion is presented of some recent experimental attempts to provide converging operations for a concept of perceptual defense. For present purposes, perceptual defense is defined as a relative failure of perception per se due to the emotional character of the stimulus. Experiments are evaluated and discussed mainly in terms of their ability to eliminate as an explanation the response-bias hypothesis for differential accuracy between neutral and emotional words. In addition, following Blum (1955), a stimulus-effect hypothesis is described and applied; all the experiments discussed require, if differences in accuracy of recognition are to be attributed to perceptual variation, converging operations to eliminate this hypothesis. It is argued as well that the search for converging operations for perceptual defense has implications for methodology in other areas of perception.

"The term 'perceptual defense' is misleading since there is no experiment known to this reviewer that contains the necessary converging operations that would permit a conclusion of defense in perception, per se," writes Eriksen (1957, p. 202). He is referring to the conceptual distinction between perceptual and response systems emphatically drawn by Garner, Hake, and Eriksen (1956), and to their general proposals for an experimental methodology to distinguish perceptual effects from response and other effects. In a recent article, Bootzin and Natsoulas (1965) claimed that the research area of perceptual defense has been exceptional in its concern with this distinction: "In fact the history of the area might be well characterized as a collective search for experimental procedures and designs to reveal the workings of perception unconfounded by other variables [p. 461]." The present article discusses some of this history in the hope that awareness of the problems involved and analysis of some recent attempts to deal with them will make a solution more likely. Such a solution will be of general importance to the study of perception. There is reason to believe that the invention of converging operations which, given positive results, would permit the conclusion that *perceptual* defense has occurred will provide a methodology for the study of numerous perceptual phenomena, a methodology permitting unambiguous conclusions about perception, per se.

The following discussion may be characterized as well in a negative way, as eschewing the strategy of reducing the substantive question of perceptual defense to an instance of the operation of more general processes, for example, perceptual readiness or set (Bruner, 1957), and then either (*a*) addressing the methodological problems involved in demonstrating whether and how set affects perception as opposed to other processes,[2] or (*b*) deriving and testing predictions from a theory of set as it affects perception. This path is avoided. The present discussion is neutral with respect to *how* the emotional character of the stimulus can affect perception in order to concentrate on *how it can be shown whether* the emotional character of the stimulus affects perception. In other words, for purposes of this review, the question of explanation or mechanism is assumed to be subsequent to the development of an adequate methodology and the collection of unambiguous results about the perceptual process itself.

---

[1] The author wishes to thank Theodore E. Parks of the University of California, Davis, for his helpful critical discussion of this article.

[2] The attempts to demonstrate the influence of explicit set on perception (as distinct from other processes) have themselves an instructive and as yet inconclusive methodological history. Some important recent chapters are Lawrence and Coles (1954), Neisser (1954), Long, Reid, and Henneman (1960), and Swets and Sewall (1961). Ralph N. Haber of the University of Rochester is preparing a theoretical analysis and review of this literature.

Of the several attempts to provide the necessary converging operations for perceptual defense that will be discussed here, only the first had been reported in the psychological literature at the time of Eriksen's (1957) summary statement. Blum (1955) presented quadrads of Blacky pictures for a fixed duration to subjects (*S*s) who had to say which picture was in each of the four positions. All *S*s knew the 11 Blacky pictures very well, but none knew that the same 4 were being presented on every one of the 48 trials, rotated in positions from trial to trial. After the experiment, all *S*s stated that they were unable to recognize any of the pictures; in effect the experimental session was for them a guessing game.

Independently of the tachistoscopic portion of the experiment, it was determined for every *S* whether he had conflict and showed repression in connection with each of the 11 pictures. Of the total of 17 *S*s, 6 (5) showed both conflict and repression for 1 or more of the 4 (7) pictures (not) presented, 14 (14) *S*s showed no conflict or repression for 1 or more of the pictures (not) presented. These are the four categories or conditions that Blum (1955) uses to argue for a perceptual effect; in the case of the presented pictures, his prediction is that there should occur fewer calls of emotional than neutral pictures and no difference should appear between emotional and neutral absent pictures. The assumptions are that, "(*a*) the subliminal stimulus would activate the conflict; and (*b*) the necessity to verbalize the conflict area at the conscious level (by calling it) would elicit defensive behavior [p. 26]." The results were as expected: the mean number of calls per picture were 9.42, 17.12, 15.20, and 16.69 for presented pictures with and without conflict and repression, and absent pictures with and without conflict and repression, respectively.

Although the purpose of Blum's experiment was to demonstrate the need for a concept of perceptual defense, his predictions follow from an assumption that defense occurs at the response level (with "the necessity to verbalize the conflict area"). It is fairly certain that what Blum means by perceptual defense is *defense instigated by a stimulus but occurring in the response system* (hereafter referred to as the "stimulus-effect hypothesis"). Consistently with this, what Blum appears to have shown is that a subliminal stimulus can affect response tendencies. He proposes that this occurs as a result of the stimulus activating a conflict which in turn affects response probabilities. This view of perceptual defense needs to be distinguished from the one implied by Eriksen's (1957) summary statement about "defense in perception, per se" and adopted for purposes of this discussion. Perceptual defense here represents relatively "a failure of the perceptual system" (Garner, Hake, & Eriksen, 1956, p. 153). Intuitively stated, the tachistoscopic stimulus word is not seen as clearly, in as great detail, etc., when it is a source of conflict or anxiety.

Blum (1955) then seems not to have provided the necessary operations converging on the concept of perceptual defense. However, he seems to have moved the search forward by his attempt to evaluate response tendencies. Thus the purpose of the absent pictures (while *S* believed these pictures were being presented and responded as such) was to provide a base line of comparison for the effects of true or appropriate stimuli on the corresponding responses. Similarly, perceptual differences will have to be demonstrated under conditions in which response tendencies on their own cannot be the source of the relevant differences in results.

Mathews and Wertheimer (1958) reported an experiment similar to Blum's entitled "A 'Pure' Measure of Perceptual Defense Uncontaminated by Response Suppression." In this experiment, an indirect attempt was made to secure a score for each *S* representing his tendency not to call the emotional words presented to him. The *S*s were given a list of eight words and told that every word flashed on the screen (singly at a fixed brief duration) would be one of them. Four of the words were neutral (by the authors' judgment) and four were selected as emotional for each *S* on the basis of a prior word-association test. The neutral and emotional words were matched for length, "approximate structure," and frequency of occurrence (Thorndike-Lorge General Word Count). As Blum (1955) had done, the experimenter (E) flashed only some of the expected materials, four of the eight words, two emotional and their matched neutrals, providing again four categories: words presented, neutral and emotional; words absent, neutral and emotional.

For every S two z scores were calculated. On the assumption of no difference between emotional and neutral words in response probabilities, one would expect for the absent words a number of calls of emotional words equal to Np where N is the total number of calls of the absent words and p by the assumption is .5. An S's z score for response suppression would be the actual number of absent, emotional words called minus the expected number (assuming no response suppression) in σ units. The same scoring procedure was used as well for the presented words; the z score in this case is believed to be a measure of both response and perceptual suppression. Thus a subtraction was required to yield a "pure" measure of perceptual defense.

Mathews and Wertheimer's positive findings for "perceptual defense" can be interpreted in the same way Blum interprets his own experiment: there is a specific stimulus effect on response tendencies, reducing the calls of the corresponding emotional words. The assignment of the difference between the z scores for present and absent emotional words to perceptual defense in the sense of poorer perception (Mathews and Wertheimer refer to "pure" and "true" perceptual defense) is faced with Blum's competing stimulus-effect hypothesis. Converging operations are needed to eliminate it.

Neither Blum nor Mathews and Wertheimer gave adequate attention to the accuracy of their Ss' performance. Differential accuracy between experimental conditions or kinds of stimuli is a necessary result of the operation of perceptual defense as here understood; if somehow avoidance or defense is occurring within his perceptual system, S would not see emotional stimuli as well as neutral ones, and therefore make more identification errors. The mere demonstration of fewer calls of emotional stimuli seems always to leave the hypothesis open: Ss are perceiving both kinds of stimuli equally well, but responding to them differently. The next experiment discussed uses both a measure of differential accuracy between anxiety-arousing and neutral words and a response-bias score for the same words.

Goldstein (1962) attempted (*a*) to demonstrate that a response bias against calling anxiety-arousing words can give the *appearance* of a perceptual-defense effect, that is, fewer correct identifications of anxiety-arousing words than of neutral control words, and (*b*) to determine whether, over and above the effect of response bias on differential accuracy, there is an effect attributable to perceptual variation. The first purpose was realized through the inclusion of a "stimulus-absent" experimental group similar to that used by Goldiamond and Hawkins (1958). On every trial, the members of this group were stimulated merely by typewritten hashmarks while being led to believe that words were being presented for subliminal durations. These Ss, as well as those presented with true stimuli (the stimulus-present group), had to choose a response from an available list of eight words, four neutral and four anxiety-arousing (as independently determined for each S). The stimulus-present group was exposed on every trial to one of the eight words, and the words were presented equally often.

The magnitude of each S's response bias was found by substracting his number of calls of neutral words from half (64) of his total number of responses. For the stimulus-absent group the mean response bias was −4.55, for the stimulus-present group −4.90, both statistically significant biases against anxiety-arousing words. Furthermore, in the case of the stimulus-absent group, it was shown that this response bias produced a mean "pseudoaccuracy" score significantly favoring the neutral words by 2.45 more words correct. Pseudoaccuracy scores were obtained by scoring, as correct or not, the responses of the stimulus-absent group in terms of correspondences with the stimulus-present group. This is clear-cut evidence that a response bias against calling anxiety-arousing words can produce what might have been accepted once as a perceptual-defense effect.

The stimulus-present group's mean differential accuracy score also favored the neutral words, by 4.50 more words correct than anxiety-arousing words. This was not significantly greater than the differential pseudoaccuracy, indicating that presentation of true stimuli did not increase the amount of defense shown. Goldstein concluded that the results support a reponse probability theory of perceptual defense; they do not display significant variation attributable to perceptual differences between anxiety-arousing and neutral words.

Goldiamond (1962) argues more generally against ascribing differences in performance between stimulus-absent and stimulus-present conditions to perception. He points out that such ascribing is based on the often unlikely assumption that *S*'s response biases are the same although different stimuli are presented. When two conditions do differ in response bias, differences between them in performance, that is, in the differential accuracy of identifying neutral and anxiety-arousing words, cannot be ascribed readily to a perceptual process. In Goldstein's experiment, the mean response biases of the two experimental groups do not appear to differ significantly; if differences in performance had been found between them, these differences presumably could have been attributed to perception, so far as the present argument has gone. However, the matter of response bias is more complicated. Goldstein, Himmelfarb, and Feder (1962) claim that the pattern of response bias may not be identical between conditions even when the tendency to choose neutral over emotional words is found to be equal: "Thus not only is it possible for the subject to show a response bias between classes of stimulus words, but it is possible for him to show a within-set bias resulting from the manner in which he distributes his guesses within a set [p. 60]."

Suppose *S* has a tendency *not* to use one of a set of anxiety-arousing words, a negative tendency which is greater than that for the anxiety-arousing set as a whole. Under stimulus-absent conditions, pseudoaccuracy would be unaffected by this so long as the bias against the total set of anxiety words remained the same. On the other hand, under stimulus-present conditions, the operation of such a bias would limit the level of performance on that particular anxiety word; despite perceiving a fragment of the stimulus ordinarily sufficient (e.g., in the case of the neutral words) to identify the word, *S* would be relatively likely to make an incorrect response. Thus, as Goldstein, Himmelfarb, and Feder (1962) put it, the "residual difference in accuracy between anxiety-words and neutral words," that is, residual after the pseudoaccuracy correction, would not "necessarily support . . . the hypothesis of a perceptual blocking mechanism [p. 60]." Within the context of Goldstein's (1962) experiment it is necessary, therefore, to demonstrate that the distribution of response probabilities is similar within neutral and anxiety-arousing sets of alternative responses.

The experiment of Bootzin and Natsoulas (1965) makes irrelevant the response bias *between* anxiety-arousing and neutral words by always pitting against each other, for response, words of the same class. Secondly, it demonstrates that the distribution of response probabilities *within* the neutral and anxiety-arousing sets is the same. Finally, with response bias eliminated as a source of variation, it shows that the accuracy of identifying anxiety-arousing words is not as great as the accuracy for neutral words. This difference is attributed to a "perceptual blocking mechanism" on the grounds that the competing, response-probability hypothesis has been eliminated.

Two quadrads; each containing two neutral and two anxiety-arousing words (as independently determined for each *S*), all words within a quadrad matched in length, first letter, and Thorndike-Lorge General Word Count; were selected for each *S*. The eight words and a card with five hashmarks were exposed tachistoscopically singly in a random order; the hashmarks always for .01 second (32 times), the words four times each for .01 second and four times each for .03 second. Following each exposure, *S* turned the page of a booklet and checked one of a pair of words on a sheet; however, he had had no information prior to a stimulus-presentation as to which words these would be. The choice was always between two neutral words or two anxiety-arousing words.

Differences associated with personality types complicate the results, and the following are the overall findings. No differences were found except at the .03-second duration where accuracy for both neutral and anxiety-arousing words was well above chance levels and there was significantly less accuracy for anxiety-arousing than for neutral words. The measure of response bias taken was the proportion of choices of the favored alternative within the pairs of anxiety-arousing or neutral words. In other words, the question was asked as to whether *S*s tended to choose as a response one of a pair of anxiety-arousing words more often than one of a pair of neutral words or vice versa. It was found that the difference in accuracy between recognition of anxiety-arousing words and neutral words was accompanied by no significant difference in response bias as thus defined.

Bootzin and Natsoulas argue that this experiment provides evidence for perceptual defense uncontaminated by response bias. They point out that the difference in accuracy (at .03-second duration) found between anxiety-arousing and neutral words cannot be attributed to differences either in response bias or to differences in the magnitude of the stimulus-word fragment affecting *S*. This leaves, according to them, perceptual variation as the likely source of the observed difference in accuracy.

The difference in accuracy between the two kinds of words, although uncontaminated by between-set and within-set response biases, could be attributed to postperceptual and preresponse events which are affected by the emotional character of the stimulus. If the difference in accuracy is due to these postperceptual events, then *S* must be seeing the anxiety-arousing words as well as he sees the neutral words, but "losing" more of the information received in the brief time prior to response. This amounts to an immediate memory or categorization hypothesis requiring, if that aspect of perception in which avoidance takes place is to be isolated, appropriate converging operations. Reference is made to an *aspect* of perception because the question of what constitutes perception as opposed to categorization or even immediate memory is problematical. Some students of perception have defined perception as a kind of categorization, others have pointed out that the central process does not terminate with the stimulus. Swets, Tanner, and Birdsall (1961) have proposed that perception is a choice among Gaussian variables, albeit a greater number than in the case of signal detection. The decision processes involved in making the choice are included in their definition of perceptual processes to distinguish them from sensory processes "on the grounds that the former must be accounted for in terms of events presumed to occur at higher centers whereas the latter can be accounted for in terms of events occurring within the receptor systems [p. 338]." Two basic intervening variables, on which depend the responses of the observer, are postulated by signal detection theory: sensitivity (d′) and the criterion used in making a judgment. As Atkinson (1963) points out, signal detection theory assumes that the sensitivity of the observer is fixed or varies randomly over time while the criterion varies in a systematic way depending on a number of variables affecting set or motivation. In its place, Atkinson (1963) proposes a variable sensitivity theory of signal detection which allows for systematic changes in sensitivity from trial to trial more nearly consistent with a conventional definition of perception, changes possibly due to "orienting responses, peripheral changes within the sensory system, or events presumed to occur at higher centers." Thus, Swets et al. (1961) include a great deal in their concept of perception—all of the postperceptual and preresponse decision processes—while Atkinson (1963) is concerned with an "activation process" distinct from the "decision process," a concept of sensitivity which may not be determined exclusively by sensory factors.

While Bootzin and Natsoulas (1965) find it necessary to demonstrate the absence in their experiment of differential within-set biases between neutral and emotional word pairs, Goldstein, Himmelfarb, and Feder (1962) seek to eliminate the relevance of all response biases to differential accuracy scores. Turning from Goldstein's (1962) statistical-correction method, they adopt a spatial forced-choice procedure. This procedure, resembling the forced-choice methods developed by Blackwell (1953), does not require the subject to verbalize or otherwise indicate what word has been presented. Instead, the subject is asked to locate spatially a particular word when it is presented along with one or more others. Goldstein, Himmelfarb, and Feder (1962) presented on every trial (at a fixed brief duration) a pair of words, one an anxiety-arousing word, the other its matched, neutral control. The task for *S* was to call out the position of one of the words as instructed by *E*, while knowing, too, what the other word would be. The kind of response bias which could be present in this situation would involve a tendency to say "right" (or "left") when a particular kind of word had to be located. Such a bias was found: most *S*s under both stimulus-present and stimulus-absent conditions tended to say "left" when the location of an anxiety-arousing word was asked for and "right" for a neutral word.

According to Goldstein, Himmelfarb, and Feder (1962):

> Although the forced choice spatial location method is not free of its own peculiar type of response bias, the counterbalancing of position of stimulus and type of stimulus word asked for by *E* prevents the response bias of word by position from affecting the total accuracy score [p. 60].

What they are referring to is the fact that if anxiety words are "recognized" better on the left because of a tendency to say "left," they will be more poorly recognized on the right to the same extent as there is an advantage for them on the left as a result of the same bias.

As proposed in discussing Goldstein's (1962) experiment, this general line of argument is valid for chance congruences, but not applicable where *S*'s responses begin to be affected by the stimulus. Consider a particular anxiety-arousing word and its matched, neutral control on two trials in both of which the anxiety-arousing word is on the right and the neutral one on the left. On the one trial the anxiety word must be located, and on the other, the neutral word. Assume further (*a*) that on the two trials the words which must be located are perceived equally well, and (*b*) the tendency to say "left" to the instruction to locate the anxiety-arousing word is stronger than the tendency to say "right" when the neutral word's position is required. Because of the relatively strong bias in the case of the anxiety-arousing word, *S* is more likely to respond erroneously to it than to the neutral word, without utilizing efficiently the input available to him. This nonperceptual advantage for the neutral word would not be compensated for on another trial (when an anxiety-word has to be located) where there is a chance congruence due to frequently saying "left." These chance congruences are "needed" to compensate for the relatively few chance congruences occurring when the anxiety word is on the right and must be located. Thus, in the context of this forced-choice procedure as well, it is necessary to demonstrate that in responding with respect to neutral and emotional words there is no greater response bias in the one case than in the other.

No difference was found by Goldstein, Himmelfarb, and Feder (1962) in the accuracies of locating neutral and anxiety-arousing words. In addition, there appears to be no difference between neutral and anxiety-arousing words with respect to the magnitude of their respective position biases. However, interpretation of the experiment meets one obstacle: *S*s' knowledge of which two words are flashed on a trial permits location of both as a result of perceiving one or part of one. Thus the perceptual defense effect may not manifest itself because partial recognition of the neutral member of a pair permits location of the anxiety-arousing member even though it is not seen as well. A procedure which would eliminate this alternative hypothesis would require selecting quadrads of matched anxiety and neutral words, and presenting a pair of anxiety words or a pair of neutral words on each trial. A correct response by process of elimination would still be possible, but depend on the perception of a word of the same type.

MacIntosh (1961) used a forced-choice procedure like the one just suggested. A total of 16 five-letter words were used, 8 judged by *E* to be emotionally unpleasant in meaning and 8 emotionally neutral. From these 16 words, 4 sets of 4 words, all words in a set emotionally unpleasant or all neutral, were selected and projected simultaneously at eight durations which ranged from chance level to a high level of accuracy in the performance they yielded. On each trial *S* was asked to locate a particular word. Each word had to be located once at each duration. The results showed no difference in accuracy of locating neutral and unpleasant words. Relating performance to ranking by *S* of the words in terms of pleasantness–unpleasantness did not yield a significant effect. As in the case of the procedures of Goldstein (1962) and Goldstein, Himmelfarb, and Feder (1962), interpretation of such results as indicative of perceptual defense (had differences been found) would depend on a *demonstration* of equal response biases for emotional and neutral words.

Goldstein, Himmelfarb, and Feder (1962), in discussing their own experiment, put forward a criticism applicable to MacIntosh (1961): "Possibly we have undermined the motivation for perceptual defense by insisting that the subject first *hear* the word from the experimenter before he *sees* it himself [p. 61]." Bootzin and Natsoulas (1965) argue similarly that with repeated exposures within a brief period, the anxiety reaction to a word may habituate. The requirement of the spatial forced-choice procedure that *S* hear an emotional word and keep it in mind may hasten this eventuality. Bootzin and Natsoulas show that their *S*s' performance on the anxiety-arousing words improved from the first to the second half of the tachistoscopic series to a point where, in the second half, the difference in accuracy between emotional and neutral words was no longer statistically significant. The neutral words, on the other hand, did not show a significant increase in accuracy between the two halves of the series of tachistoscopic exposures.

Another experimenter who has attempted to demonstrate a perceptual effect is Zajonc (1962), whose procedure, considerably different from those already reviewed, involves three consecutive phases: (*a*) determination of visual duration thresholds for recognition of 6 taboo words and 6 neutral words by the ascending method of limits, with GSR readings taken on every presentation of a stimulus word; (*b*) a paired-associates learning task in which these 12 words were the stimuli and an additional 12 (6 taboo and 6 neutral) served as responses; (*c*) a second determination of thresholds and GSRs for the stimulus words in which half the *S*s had to respond with the response words learned in the paired associate task and half had to respond in the usual manner, by calling the stimulus word.

The point of the experiment was to distinguish the contribution to threshold measures and GSR measures of the response made as opposed to the stimulus presented. The first part of the experiment (phase *a* above) would not discriminate between the stimulus and response effects because a taboo word had to be called when recognized. Zajonc found that the taboo words provided higher recognition thresholds and greater GSRs. Garner, Hake, and Eriksen (1956) propose two hypotheses to deal with such results: either "the perceptual system differentially discriminates on the basis of the emotional content of the perceived stimuli" or "the difference in threshold is a result of a characteristic of the response system, which inhibits the verbalizing of some of the words [p. 151]." They then suggest that the next steps taken by Zajonc in his procedure would be the necessary converging operations to eliminate one or the other of the two hypotheses.

The third phase of the procedure, by requiring *S*s to call at times words from the neutral category to indicate perception of taboo words and of other neutral words, and to call taboo words to indicate perception of neutral words and of other taboo words (associates learned in phase *b*), permits evaluation of the contribution of the stimulus and of the response to the measured thresholds. For the group responding with the learned associates, taboo and neutral *stimuli* did not differ significantly in recognition thresholds although *responding* with taboo words yielded higher recognition thresholds than responding with neutral words. The other experimental group provided recognition-threshold results similar to those in phase *a*. Some effect of the stimulus was shown on GSR, but the response was the primary determinant. From these results Zajonc (1962) concludes, "The evidence presented failed to disclose perceptual effects of any significance. The recognition threshold was found to be a function not of what *S* saw but what he had to say [p. 213]."

If the *S*s had not seen the taboo words as well as they saw the neutral ones, there would certainly have been an observed stimulus effect. To evaluate the design with regard to its ability to demonstrate perceptual defense, it is necessary to ask whether the occurrence of an observed stimulus effect would have provided unambiguous evidence for a *perceptual* effect (in the sense indicated earlier of a relative failure *in* perception). Blum's (1955) stimulus-effect hypothesis would seem to be a competing hypothesis requiring elimination by converging operations; in Zajonc's experiment an observed stimulus effect could be interpreted in terms of this hypothesis by stating that taboo words produce conflict which inhibits the appropriate responses from occurring. The fact that the response to a taboo stimulus was a neutral word would not eliminate Blum's hypothesis, because the inhibition may be extended to responses other than the reading response associated with the stimulus.

Zajonc (1962) views "defensive perceptual blocking" as equivalent to the "contribution of the stimulus [p. 206]," but Garner, Hake, and Eriksen's (1956) reference to the perceptual system's differentially discriminating suggests that they are adopting a definition of perceptual defense as occurring within the perceptual system. The next step in the search for the necessary converging operations for perceptual defense will require, if possible, the elimination of the stimulus-effect hypothesis, just as recent experiments have been designed to eliminate the hypothesis of the response effect. The experiments of Goldstein (1962), Bootzin and Natsoulas (1965), Goldstein, Himmelfarb, and Feder (1962), and MacIntosh (1961), not previously discussed in relation to the stimulus effect, can all be interpreted according to Blum's hypothesis. For example, the Bootzin and Natsoulas findings of greater accuracy for neutral than emotional words were interpreted as showing perceptual differences, because the response effect was shown not to be responsible for the difference. However, the difference in accuracy may be due to *S*'s perceiving both kinds of stimuli to the same degree, but responding differently as a result of the conflict engendered by the emotional words. This would not appear in the response bias scores because it would work both ways: when emotional word *a* was flashed there would be a tendency to avoid calling it, as there would be when emotional word *b* of the pair was flashed. The response bias score refers to the preferred response, but neither *a* nor *b* would be preferred more than the matched neutral words, unless they differed significantly in the degree of conflict they produced.

A last point concerns the assertion made at the beginning that a search for converging operations for perceptual defense will be repaid by the discovery of methods having more general applicability in the study of perception. Again to illustrate, the Bootzin and Natsoulas (1965) procedure may be considered. The stimulus effect which must be eliminated as an alternative hypothesis if true perceptual defense is to be demonstrated is not relevant in situations where emotionality or conflict is not a variable. In such cases, to find that stimuli which vary in a systematic way and are projected for equal durations yield different accuracies of identification, where between-set response biases are irrelevant and within-set biases are demonstrably equal, is to find evidence for perceptual variation. Taylor, Rosenfeldt, and Schulz (1961) would object to this conclusion on the grounds of an alternative hypothesis to perceptual variation. According to this hypothesis, two classes of stimuli which are shown to differ in how accurately they are identified with a forced-choice technique such as that of Bootzin and Natsoulas (1965) might so differ because of the relation of perception to response. Perception may be constant across conditions but guessing or identifying on the basis of partial perception may differ between them. Taylor, Rosenfeldt, and Schulz (1961) suggest, for purposes of eliminating or reducing the likelihood of this competing hypothesis,

> If a variable that has been shown to influence threshold behavior can be shown *not* to influence response identification when partial cues are presented at a suprathreshold level, a firmer basis would be provided to support a perceptual interpretation of the former [p. 495].

In a forced-choice procedure of the kind used by Bootzin and Natsoulas (1965), it seems likely, although it must be shown, that the ready availability of the responses to *S* would prevent this kind of variation from affecting differential accuracy scores.

No position has been taken, in the above methodologically directed discussion, with respect to *how* the emotional character of the stimulus may have its negative effects upon the perceptual process itself. Such considerations are left to the development of psychological and perhaps neurophysiological theory; therefore, the present view is neutral as to whether, for example, it is scientifically useful to introduce such a concept as "unconscious perception," in order to explain the relative failure of perception, per se. Similarly, the stimulus-effect hypothesis must not be construed as involving necessarily the proposal of a specific mechanism. All that this hypothesis maintains is that stimuli of a certain emotional character can have specific effects of defense or avoidance in the response system, as a result of the conflict or emotion they produce, while leaving the perceptual system unperturbed.

REFERENCES

Atkinson, R. C. A variable sensitivity theory of signal detection. *Psychological Review,* 1963, **70,** 91–106.

Blackwell, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *University of Michigan Engineering Research Institute Bulletin,* 1953, No. 36.

Blum, G. S. Perceptual defense revisited. *Journal of Abnormal and Social Psychology,* 1955, **51,** 24–29.

Bootzin, R. R., & Natsoulas, T. Evidence for perceptual defense uncontaminated by response bias. *Journal of Personality and Social Psychology,* 1965, **1,** 461–468.

Bruner, J. S. On perceptual readiness. *Psychological Review,* 1957, **64,** 123–152.

Eriksen, C. W. Personality. *Annual Review of Psychology,* 1957, **8,** 185–210.

Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. *Psychological Review,* 1956, **63,** 149–159.

Goldiamond, I. Perception. In A. Bachrach (Ed.), *Experimental foundations of clinical psychology.* New York: Basic Books, 1962.

Goldiamond, I., & Hawkins, W. F. Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *Journal of Experimental Psychology,* 1958, **56,** 457–463.

Goldstein, M. J. A test of the response probability theory of perceptual defense. *Journal of Experimental Psychology,* 1962, **63,** 23–28.

Goldstein, M. J., Himmelfarb, S., & Feder, W. A further study of the relationship between response bias and perceptual defense. *Journal of Abnormal and Social Psychology,* 1962, **64,** 56–62.

Lawrence, D. H., & Coles, G. R. Accuracy of recognition with alternatives before and after the stimulus. *Journal of Experimental Psychology,* 1954, **47,** 208–214.

Long, E. R., Reid, L. S., & Henneman, R. H. An experimental analysis of set: Variables influencing the identification of ambiguous, visual stimulus objects. *American Journal of Psychology,* 1960, **73,** 553–562.

MacIntosh, S. P. Perceptibility of emotional and nonemotional stimuli with a forced-choice method. *Dissertation Abstracts,* 1961, **21,** 2784–2785. (Abstract)

Mathews, A., & Wertheimer, M. A "pure" measure of perceptual defense uncontaminated by response suppression. *Journal of Abnormal and Social Psychology,* 1958, **57,** 373–376.

Neisser, U. An experimental distinction between perceptual process and verbal response. *Journal of Experimental Psychology,* 1954, **47,** 399–402.

Swets, J. A., & Sewall, S. T. Stimulus vs. response uncertainty in recognition. *Journal of the Acoustical Society of America,* 1961, **33,** 1586–1592.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. *Psychological Review,* 1961, **68,** 301–340.

Taylor, Janet A., Rosenfeldt, Doris C., & Schulz, R. W. The relationship between word frequency and perceptibility with a forced-choice technique. *Journal of Abnormal and Social Psychology,* 1961, **62,** 491–496.

Zajonc, R. B. Response suppression in perceptual defense. *Journal of Experimental Psychology,* 1962, **64,** 206–214.

(Received October 28, 1964)

Psychological Bulletin
1965, Vol. 64, No. 6, 402–412

CONCEPTUAL ORIENTATIONS OF CONTEMPORARY SMALL GROUP THEORY[1]

JOHN DELAMATER,[2] CHARLES G. McCLINTOCK
University of California, Santa Barbara
AND GORDON BECKER[3]
General Electric Company, Santa Barbara, California

In an attempt to study some of the conceptual and content parameters of contemporary small group orientations, 2 classificatory systems were developed and applied to a sample of empirical hypotheses derived from 6 such orientations. The results of this analysis provide data on the similarities and differences between these points of view on 3 dimensions: (a) the size of the social unit(s), (b) the social process level(s) with which they deal, and (c) the substantive content of the variables which they employ.

In surveying contemporary conceptualizations of small group behavior, one emerges with the feeling that there are both similarities and differences among them. In some places, two or more authors seem to be writing about the same variables in the same way; more importantly, in other instances they seem to view similar processes from different points of view. In any attempt to evaluate findings concerning behavior in small groups a difficulty arises in that points of agreement and disagreement are often not explicit, but rather are obscured by differences in terminology, units of analysis, levels of aggregation (cf. McClintock, Becker, & DeLamater, 1963), etc.

These similarities and differences in conceptualization are important because they are reflected in many of the critical hypotheses found in those studies which form the empirical basis of the field of small group behavior. Since many of these studies derive their conceptual orientation from one of several principal points of view, an "objectification" of differences in these approaches can illuminate considerably the underlying similarities and differences in hypotheses, techniques of operationalization, and even results extant in contemporary small group research.

The present paper presents a first attempt at such "objectification." It was felt that some understanding of theoretical similarities and differences could be gained by examining certain conceptual and content parameters of current research based on each of six principal orientations to small group behavior: Bales (1950), Cartwright and Zander (1960), Homans (1950, 1961), Moreno (1956), Newcomb (1961), and Thibaut and Kelley (1961). (For a recent review of the conceptual orientation of social psychology in general, see Lambert, 1963.) More specifically, an analysis was made of (*a*) the size of the action units around which the variables employed by these points of view are conceptualized, (*b*) the types of social processes conceptualized, and (*c*) the substantive content of the variables of the orientations as they appear in empirical studies, especially those studies conducted or cited by the theorists as illustrative of their model.

In some ways this analysis can be viewed as a complementary effort to the extensive research programs conducted by McGrath (1962a, 1962b, 1962c) and Altman and Terauds (1960). The latter have developed a method for classifying small group variables

> in terms of their operational properties rather than their conceptual/verbal properties . . . [and] not in terms of their substance or content (e.g., cohesion, group structure, etc.), but in terms of the formal data characteristics of variables, i.e., their operational definitions.

---

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract Number AF 49(638)-794.

[2] Now at the University of Michigan.

[3] Now at Yeshiva University.

Thus, the present analysis provides a measure of the current conceptual orientations in the field of small group behavior on several dimensions, that is, the extent to which these orientations cover the range of action units and the degree to which they include various types of social processes and substantive variables in their formulations. In addition, it indicates the similarities and differences between points of view on these dimensions, and thus hopefully increases our conceptual understanding of both the orientations and the research which are derived from them.

Data Selection

A sample of empirical studies representing each of the six orientations cited above was selected on the basis of four criteria. The first and primary criterion for study selection was that it be referenced in the principal work or works written about the conceptual system.[4] Empirical studies which were discussed or reprinted in these works were chosen as source studies. In each case where the major work merely discussed a study, the original article was used as the basis for analysis. Where an insufficient number of experiments were represented in the principal work, additional experimental reports were selected from those referenced in the initial set of experimental studies. The specific studies included in the analysis are indicated in Table 1.

A second criterion of selection was that all studies included in the analysis contain (*a*) a specific hypothesis or hypotheses deductively derived from or closely related to the general orientation, (*b*) a clear specification of the conceptualization of and the methods of operationalizing each variable, and (*c*) a statement and statistical analysis of the results. The presence of these characteristics was judged necessary to facilitate variable classification.

TABLE 1

Source Studies of Variables Employed in the Analysis

Conceptual orientation	Source studies
Bales	Bales, 1955, 1956; Bales & Strodtbeck, 1960; Bales & Slater, 1955; Borgatta & Bales, 1955a, 1955b; Heinicke & Bales, 1953; Talland, 1955.
Cartwright & Zander	Cohen, 1959; Deutsch & Gerard, 1960; Festinger & Thibaut, 1951; Horwitz, 1960; Hurwitz, Zander, & Hymovitch, 1960; Pepitone & Kleiner, 1957; Pepitone & Reichling, 1960; Raven & Reitsema, 1960; Schachter, Ellertson, McBride, & Gregory, 1960; Stotland & Zander, 1958.
Homans	Homans, 1950, 1961; Reicken & Homans, 1954.
Moreno	Deutschberger, 1947; Kidd, 1960; Moreno, 1960; Mouton, Bell, & Blake, 1960; Northway & Detweiler, 1956; Northway & Rooks, 1956; Tagiuri, 1952; Zeleny, 1956.
Newcomb	Newcomb, 1961.
Thibaut & Kelley	Adams, 1953; Azrin & Lindsley, 1956; Back, 1951; Blau, 1954; Cohen, 1958; Dittes & Kelley, 1956; Hoffman, Festinger, & Lawrence, 1954; Kelley & Lamb, 1957; Kelley & Volkart, 1952; Kounin, Polansky, Biddle, & Finn, 1956; Lipset, Trow, & Coleman, 1956; Mills, 1953; Spector, 1956; Thibaut & Coules, 1952; Thibaut & Riecken, 1955.

The third principle of selection was that approximately 15 tested relationships (hypotheses) be selected from each conceptual position. (The only point of view for which the optimum number of relationships was not obtained was Moreno's, due to a lack of access to studies relevant to it.) An attempt was also made to include only one tested relationship from each source study. However, where only a small number of studies were available, more than one relationship was sometimes included; this was necessary in 11 of the 55 studies with particular concentration in the orientations of Cartwright and Zander, and Moreno. It should also be pointed out that the only source of experiments based on Newcomb's approach was his recent book (1961); however, the hypotheses in this

---

[4] The primary sources for studies were as follows: Bales (Hare, Borgatta, & Bales, 1955), Cartwright and Zander (1960), Homans (1950, 1961; Riecken & Homans, 1954), Moreno (1956, 1960), Newcomb (1961), and Thibaut and Kelley (1961).

TABLE 2

Classification of Variables in Hypotheses by Size of Social Unit

Independent variable	Dependent variable			
	Individual	Dyad	Group	Total
Individual	23	1	4	28
Dyad	3	10	0	13
Group	14	1	27	42
Total	40	12	31	83

study cover all major parts of his conceptual system. Finally, only those hypotheses which were statistically supported ($p < .05$) by the experiment were included in the analysis.

In the total sample of selected hypotheses, 4 variables were unclassifiable, and thus 4 relationships (8 variables) do not appear in the subsequent tables. In addition, 11 other variables do not appear because they were minor and/or control variables in a multiple-variable hypothesis. Thus the total number of relationships classified and included in the analysis equals 83; the total number of variables initially studied was 185 (95 independent, 90 dependent).

Analysis by Size of Social Unit

The classification of variables was conducted in two phases. In the first analysis, each major variable of a given hypothesis was classified according to the size of the acting social unit. A variable, by definition, must be capable of taking on different values, and thus represents a class or dimension of some kind. In the present analysis, if the value of a variable could be determined by observing only one individual, the variable was classified at the "individual level"; if it was necessary to observe exactly two individuals to determine the value, it was classified at the "dyad level"; and if three or more individuals had to be observed, the variable was classified at the "group level." Thus, even if a variable is represented by the mean, sum, or any other characterization of a set of scores, the variable was called: (*a*) *individual*, if all the scores were obtained from a single individual, for example, an individual's score on an attitude scale, (*b*) *dyad*, if all the scores were obtained from two individuals, for example, degree of similarity of A's and B's attitudes toward X, and (*c*) *group*, if all the scores were obtained from three or more individuals, for example, the attraction of the members of the group for the group.

The results of this classification are presented in Table 2. These data show the unit-size by unit-size relationships for the total sample of hypotheses; thus, the frequency in each cell indicates the total number of hypotheses which related an independent variable at a given unit level to a dependent variable at a given level. The diagonal cells in the table provide an indication of the number of hypotheses which conceptually employ a social unit of the same magnitude in the independent and dependent variables. Table 3 presents the distribution of size of social unit for each variable in each orientation, regardless of whether they were employed in independent or dependent variables.

The results indicate that the independent and dependent variables of a given hypothesis tend to employ an acting social unit of the same size (i.e., those hypotheses which fall

TABLE 3

Classification of Variables by Orientation and Size of Social Unit

Unit size	Orientation						
	Bales	Cartwright & Zander	Homans	Moreno	Newcomb	Thibaut & Kelley	Total
Individual	3	18	10	10	9	16	66
Dyad	2	0	7	1	12	3	25
Group	27	14	11	9	5	9	75
Total	32	32	28	20	26	28	166

in the diagonal cells of Table 2); this is true in 72% of the hypotheses studied. It is also evident that the hypotheses analyzed primarily deal with social units at the individual and group levels in both independent (70 out of 83, or 84%) and dependent variables (85%). In addition, there is a greater tendency to employ the group as the social unit in independent variables than in dependent ones (57%), and to deal with the individual unit in dependent variables rather than in independent ones (60%). Dyads are employed about equally often in both types of variables.

Table 3 indicates even more clearly the tendency to use variables dealing with individuals and groups; 85% of the variables studied employ these two units. Turning to the distribution of units by orientations, three conceptual systems give approximately equal emphasis to individual and group units: Cartwright and Zander, Homans, and Moreno. The other three each tend to emphasize variables dealing with only one of the three units. Thus, a high proportion of the variables derived from Bales' approach deal with groups (84%); variables from Thibaut and Kelley's orientation tend to refer to individuals (57%); and Newcomb deals slightly more often with dyadic variables than with individual or group ones (46%).

ANALYSIS BY TYPE OF SOCIAL PROCESS AND SUBSTANTIVE CONTENT

The current controversies over conceptual and substantive orientations in social psychology involve issues concerning both the appropriateness of conceptualizing theory around various social processes and of employing various substantive variables. It should be noted that the present classificatory scheme was not designed to resolve the controversies nor to reflect a particular point of view, but rather to provide a description of the positions implied by the six conceptual schemes selected for investigation. For example, the fact that the present system permits affective and cognitive states to be attributed to either individuals or groups was not meant to imply that the authors necessarily consider either usage appropriate or inappropriate. (For a discussion of the problem of personification in the social sciences, see Allport, 1962, and Nagel, 1961.)

Social Process Categories

Individual Level. At this level of social process the individual is treated as the unit of conceptualization; there is no specification of any relationship between a given individual and other individuals or a group. This level includes variables conceptualized as states, properties, and behaviors of the individual; for example, hunger, weight, IQ, anxiety, etc.

Relational Level. Here, a one-way relationship is specified between a given individual and one or more other individuals or a group, for example, A's love for B, A's influence on B, A's status in relation to B; where A is an individual and B is another individual or group. This level includes variables dealing with the type and nature of such relationships.

Mutual Level. At this level a mutual relationship or interdependency is specified between an individual and other individual(s). This level includes such variables as the number of (dis)agreements between A and B, mutuality of attraction between A and B, intensity of conflict between A and B; where A is an individual and B is another individual or group.

Group Level. Here, the group is specified as the unit of conceptualization. This level includes variables conceptualized as states, properties, and behaviors of the group; that is, variables that deal with the group as an entity, such as the cohesiveness of A, the size of A, the productivity of A. It also includes those which deal with one part of the group relative to another.

One can also describe group relational and group mutual levels, that is, the relationship of group A to group B, or between groups A and B. However, such conceptualizations of group variables did not occur among the sample of hypotheses selected for the present study.

Substantive Content Categories

Affect. This type includes all variables which are defined and/or conceptualized as

TABLE 4

CLASSIFICATION OF VARIABLES IN HYPOTHESES BY ORIENTATION, CONCEPTUAL LEVEL, AND CONTENT TYPE

Independent variable	Dependent variable															Totals
	Individual			Relational				Mutual				Group				
	Aff.	Cog.	Beh.	Aff.	Cog.	Beh.	Pos.	Aff.	Cog.	Beh.	Pos.	Aff.	Cog.	Beh.	Struc.	
Individual Affect			TTT													3
Cognition	CCCC	T			C											6
Behavior			M ┌9				MM ┌3				┌0				┌0	3 (12)
Relational Affect	HH				NNNN	HHT										9
Cognition					C			M						C		3
Behavior				H			M									2
Position		CC	HM ┌6		MT	HTT	┌15				┌1	B		H	H ┌4	12 (26)
Mutual Affect									T	T					NM	4
Cognition								NN								2
Behavior								H	H	B						3
Position			┌0		N		┌1	NNN			┌10				NN ┌4	6 (15)
Group Affect			BC	T									C	CT		6
Cognition						H	BM					BC	B	B	H	8
Behavior												B	CH			3
Structure		B	BMT ┌6		C		┌5				┌0	T		BBBBB BC	┌19	13 (30)
Totals	6	4	11 (21)	2	10	7	5 (24)	7	2	2	0 (11)	5	4	12	6 (27)	83

Note.—B = Bales, C = Cartwright & Zander, H = Homans, M = Moreno, N = Newcomb, T = Thibaut & Kelley.

internal affective states, namely, emotional feelings, both positive and negative. The source of such states can be at any level of conceptualization; for example, John is unhappy, John does(not) like Harry, John and Harry do(not) like each other, the group does(not) like Harry. Motivational variables are included in this category.

Cognition. This type includes all variables which are conceptualized as internal cognitive states or processes of an individual or a group. Variables dealing with internal beliefs and perceptions are of this type. Cognitive variables may be at any level of conceptualization; for example, Harry believes . . ., the group thinks . . . .

Behavior. This type includes variables which refer to the overt behavioral manifestations of individuals or groups. It includes all observable motor acts including verbalizations and direct expressions of affect and cognitions except for introspective reports. Such variables may be at any level of conceptualization; for example, Harry runs, the group convenes.

Position. This type includes all variables which are attributes of an individual or a group which are defined only in terms of the individual's or group's relationship to other individuals or groups. It includes variables such as the individual's or the group's role, leadership, status, sociometric position, etc. (In the variables analyzed, this type did not occur at the group level.)

Structure. This type includes all variables which deal with the number, location, and distribution of parts in a unit. It can be at any level; for example, the arrangement of an individual's nerve cells, the size of a group, the distribution of roles in a group, etc. (In the variables analyzed, however, this type occurred only at the group level.)

RESULTS

The results of the classification of the selected hypotheses in terms of the social process level and content type are presented in Table 4. Each cell entry stands for one hypothesis which related an independent variable of the indicated process level and content type to a dependent variable of the

TABLE 5

CLASSIFICATION OF VARIABLES BY ORIENTATION AND LEVEL OF SOCIAL PROCESS

Level	Orientation																				
	Bales			Cartwright & Zander			Homans			Moreno			Newcomb			Thibaut & Kelley			Totals		
	Ind.	Dep.	Tot.	Ind.	Dep.	Tot.	Ind.	Dep.	Tot.	Ind.	Dep.	Tot.	Ind.	Dep.	Tot.	Ind.	Dep.	Tot.	Ind.	Dep.	Both
Individual	0	3	3	5	7	12	0	3	3	3	3	6	0	0	0	4	5	9	12	21	33
Relational	1	1	2	4	3	7	9	5	14	4	5	9	4	5	9	4	5	9	26	24	50
Mutual	1	1	2	0	0	0	2	2	4	1	1	2	9	5	14	2	2	4	15	11	26
Group	14	11	25	7	6	13	3	4	7	2	1	3	0	3	3	4	2	6	30	27	57
Totals	16	16	32	16	16	32	14	14	28	10	10	20	13	13	26	14	14	28	83	83	166

TABLE 6

CONTENT CATEGORIES OF INDEPENDENT AND DEPENDENT VARIABLES BY ORIENTATION

Orientation	Content category															Total sample size
	Affect			Cognition			Behavior			Position			Structure			
	I.V.	D.V.	%	I.V.	D.V.	%	I.V.	D.V.	%	I.V.	D.V.	%	I.V.	D.V.	%	
Bales	1	3	12	4	2	19	2	10	38	1	1	6	8	0	25	32
Cartwright & Zander	3	5	25	8	7	47	1	4	16	2	0	6	2	0	6	32
Homans	4	4	29	2	2	14	4	6	36	4	0	14	0	2	7	28
Moreno	1	1	10	2	1	15	4	3	35	2	4	30	1	1	10	20
Newcomb	5	5	38	2	5	27	0	0	0	6	0	23	0	3	12	26
Thibaut & Kelley	8	2	36	1	3	14	0	9	32	3	0	11	2	0	7	28
Total	22	20	25	19	20	24	11	32	26	18	5	14	13	6	11	166

Note.—Percent figures indicate the percentage of the total number of variables from a given orientation (or of the total number classified) which fall in the indicated content category.

indicated level and type. The letter indicates the conceptual framework from which the hypothesis was derived. The number in the lower right-hand corner of each level-level cell indicates the number of hypotheses which related those two process levels. Marginal row (independent) and column (dependent) totals indicate the number of variables found for each process level and content type, summed over all orientations.

The results indicate that independent and dependent variables tend to be conceptualized at the same social process level; independent variables at the individual level tend to be related to dependent variables at the individual level, and mutual to mutual. Thus, 64% of the hypotheses analyzed fall in the diagonal cells of the matrix of Table 4. This predisposition to conceptualize the independent and dependent variables of a given hypothesized relationship at the same social process level characterizes a majority of the hypotheses selected from each orientation except Moreno's.

Table 5 presents a classification of the variables by orientation and by level of social process. The data indicate that the greatest proportion of the variables are conceptualized at the group (34%) and the relational (30%) levels. A greater proportion of variables conceptualized at the individual level are treated as dependent variables (64%) and a greater proportion of mutual level variables are treated as independent variables (58%).

Table 5 also indicates that hypotheses derived from a given conceptual system tend to emphasize only one or two levels of social process. Thus, 78% of the variables derived from Bales' approach are conceptualized at the group level, 78% of those from Cartwright and Zander's are at the individual and group levels, 75% from Homans' are at the relational and group level, 75% from Moreno's are at the relational and individual level, and 88% from Newcomb's are at the relational and mutual level. This tendency is less marked in the variables derived from Thibaut and Kelley's system; 64% of these variables are at the individual and relational levels.

Table 6 shows the use of content categories across orientations and social process levels. The results indicate that variables conceptualized in terms of overt behavior are more frequently employed as dependent variables (74%) than as independent variables, whereas a smaller proportion of position (22%) and structure (32%) variables are conceptualized as dependent variables. Affect and cognition are treated equally often as independent and dependent variables.

Table 6 further indicates that different conceptual systems tend to emphasize different types of content. Note that 63% of Bales' variables fall within the behavior and structure categories, 72% of Cartwright and Zander's fall within affect and cognition, 65% of Homans' within affect and behavior, 65% of Moreno's within behavior and position, 65% of Newcomb's within affect and cognition, and 68% of Thibaut and Kelley's within affect and behavior.

DISCUSSION

In reviewing these six major orientations to the study of small group behavior, considerable heterogeneity was discovered in the conceptual approaches. The variables employed in representative hypotheses differed as to the size of the acting unit, the conceptual level of social process, and the type of substantive content emphasized. Despite this heterogeneity, there do exist some common tendencies. The results indicated a tendency among all the hypotheses classified to use the same sized social units (72%) for the independent and dependent variable of a given hypothesis. The independent and dependent levels of social process were also found to be at similar levels (64%). Furthermore, there was a tendency to employ the individual more often as the acting social unit (60%) and as the level of social process (64%) on the dependent side of a hypothesized relationship. The latter predisposition is consistent with the historic orientation of social psychology towards viewing an individual's behavior as dependent upon given social characteristics of the environment.

In comparing the conceptual orientations of individual approaches, there are a number of findings which are more or less consistent with contemporary characterizations of them. In terms of the level of social process, one observes that the variables selected as representative of Bales' approach are conceptualized primarily at the group process level which is consistent with a more sociological orientation, whereas the variables of Cartwright and Zander's are primarily conceptualized at either the group or individual level. The latter is congruent with Lewin's (1951) emphasis upon the social behavior of both the individual and the group, and a lack of theoretical concern about the nature of the social process (relational or mutual) between individuals. In contrast, the remaining orientations (Homans, Moreno, Newcomb, Thibaut, and Kelley) place rather heavy emphasis upon the dynamics of the interpersonal relationship per se. It should be noted that despite repeated assertions by social psychologists that the most critical area for investigation in small group behavior is the reciprocity of the interaction process, only those hypotheses stemming from Newcomb's orientation had a substantial number of variables conceptualized at the mutual level of social process.

In the selection of substantive variables, those hypotheses which derive from the more sociological points of view (Bales and Moreno) were found to select more variables referring to position, structure, and overt behavior than the other orientations. This emphasis may in part be due to the fact that it is more appropriate in sociology to talk about various characteristics of groups and their behavior rather than affect or cognition. In both Homans and Thibaut and Kelley one observes the heaviest stress upon the substantive variables of affect and behavior which is consistent with their reward-cost (reinforcement) formulations of small group behavior. Finally, the substantive variables most characteristic of both Newcomb's and Cartwright and Zander's conceptual orientations were those of affect and cognition which reflects their stress upon the importance of attitudinal and perceptual variables in social behavior.

It is of interest to note in the present analysis those levels of social process which are not related frequently in the present set of sampled hypotheses. Thus, there are few, if any, hypotheses which utilize variables at the individual process level as independent, and mutual level variables as dependent variables and vice versa. A study of research investigations stemming from more clinically oriented personality theories would probably reveal a substantial number of hypotheses at these levels. Thus, this would appear to be an area in which greater rapprochement between the two specialties might be achieved.

Furthermore, there are relatively few occasions when variables at the individual process level are treated as the independent and at the group level are treated as the dependent—that is, what characteristics of the individual affect the group? As noted previously, there are strong historical precedents to state hypotheses in the opposite direction, namely, from group to individual level. Although a more powerful explanatory model could undoubtedly be achieved by employing relational or mutual variables as independent variables when the group is treated as a dependent variable, it is of interest to note the continuous prominence which the individual is afforded in past and contemporary social psychological theory.

In conclusion, it should be emphasized that the present analysis was based upon a sample of empirical studies selected to represent various conceptual approaches to the study of small group behavior. The results of this analysis are limited by the sample chosen, although the authors feel that this sample was adequate and representative. In addition, there is often a discrepancy between the comprehensiveness of the formulations of a given conceptualization and the range of empirical studies deriving from that orientation. Thus, most orientations in social psychology conceptually deal with mutual relationships more frequently than would appear to be the case by such an examination of the empirical literature. The lack of research focused upon the interdependencies or mutualities of interpersonal behavior may result from an inability to state precisely the conceptual nature of these relationships or from present methodological difficulties in measuring reciprocal behavior. Finally, it should be noted that conceptualization often anticipates research; a sample of empirically-tested propositions may thus be somewhat deficient in representing the content of a conceptual system.

REFERENCES

Adams, S. Status congruency as a variable in small group performance. *Social Forces,* 1953, **32,** 16–22.

Allport, F. H. A structuronomic conception of behavior: Individual and collective. I. Structural theory and the master problem of social psychology. *Journal of Abnormal and Social Psychology,* 1962, **64,** 3–30.

Altman, I. C., Pendleton, C., & Terauds, A. *Annotations of small group research studies.* (HSR-RR-60/5-Gn) McLean, Va.: Human Sciences Research, 1960.

Altman, I. C., & Terauds, A. *Major variables of the small group field.* (HSR-RR-60/6-Gn) McLean, Va.: Human Sciences Research, 1960.

Azrin, N., & Lindsley, O. R. The reinforcement of cooperation between children. *Journal of Abnormal and Social Psychology,* 1956, **52,** 100–102.

Back, K. W. Influence through social communication. *Journal of Abnormal and Social Psychology,* 1951, **46,** 9–23.

Bales, R. F. *Interaction process analysis: A method for the study of small groups.* Cambridge, Mass.: Addison-Wesley, 1950.

Bales, R. F. The equilibrium problem in small groups. In A. P. Hare, E. F. Borgatta, & R. F. Bales (Eds.), *Small groups: Studies in social interaction.* New York: Knopf, 1955. Pp. 424–456.

Bales, R. F. Task status and likeability as a function of talking and listening in decision-making groups. In L. D. White (Ed.), *The state of the social sciences.* Chicago: Univer. Chicago Press, 1956. Pp. 148–161.

Bales, R. F., & Slater, P. E. Role differentiation in small decision-making groups. In T. Parsons et al. (Eds.), *Family, socialization and interaction process.* Glencoe, Ill.: Free Press, 1955. Pp. 259–306.

Bales, R. F., & Strodtbeck, F. L. Phases in group problem solving. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 624–638.

Blau, P. M. Patterns of interaction among a group of officials in a government agency. *Human Relations,* 1954, **7,** 337–348.

Borgatta, E. F., & Bales, R. F. Interaction of individuals in reconstituted groups. In A. P. Hare et al. (Eds.), *Small groups: Studies in social interaction.* New York: Knopf, 1955. Pp. 379–396. (a)

Borgatta, E. F., & Bales, R. F. Size of group as a factor in the interaction profile. In A. P. Hare et al. (Eds.), *Small groups: Studies in social interaction.* New York: Knopf, 1955. Pp. 396–413. (b)

Cartwright, D., & Zander, A. (Eds.) *Group dynamics: Research and theory.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.

Cohen, A. R. Upward communication in experimentally created hierarchies. *Human Relations,* 1958, **11,** 41–53.

Cohen, A. R. Situational structure, self-esteem, and threat-oriented reactions to power. In D. Cartwright (Ed.), *Studies in social power.* Ann Arbor: Univer. Michigan Press, 1955. Pp. 36–52.

Deutsch, M., & Gerard, H. B. A study of normative and informational social influences upon individual judgment. In D. Cartwright & A. Zander (Eds), *Group dynamics: Research and theory.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 201–213.

Deutschberger, P. The *tele*-factor: Horizon and awareness. *Sociometry,* 1947, **10,** 242–249.

Dittes, J. E., & Kelley, H. H. Effects of different conditions of acceptance upon conformity to group norms. *Journal of Abnormal and Social Psychology,* 1956, **53**, 100–107.

Festinger, L., & Thibaut, J. Interpersonal communication in small groups. *Journal of Abnormal and Social Psychology,* 1951, **46**, 92–99.

Hare, P., Borgatta, E. F., & Bales, R. F. (Eds.) *Small groups: Studies in social interaction.* New York: Knopf, 1955.

Heinicke, C., & Bales, R. F. Developmental trends in the structure of small groups. *Sociometry,* 1953, **16**, 7–38.

Hoffman, P. J., Festinger, L., & Lawrence, D. H. Tendencies toward group comparability in competitive bargaining. *Human Relations,* 1954, **7**, 141–159.

Homans, G. C. *The human group.* New York: Harcourt, Brace, 1950.

Homans, G. C. *Social behavior: Its elementary forms.* New York: Harcourt, Brace & World, 1961.

Horwitz, M. The recall of interrupted group tasks. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 370–394.

Hurwitz, J. I., Zander, A. F., & Hymovitch, B. Some effects of power on the relations among group members. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory.* (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 800–809.

Kelley, H. H., & Lamb, T. W. Certainty of judgment and resistance to social influence. *Journal of Abnormal and Social Psychology,* 1957, **55**, 137–139.

Kelley, H. H., & Volkart, E. H. The resistance to change of group-anchored attitudes. *American Sociological Review,* 1952, **17**, 453–465.

Kidd, J. W. An analysis of social rejection in a college men's residence hall. In J. L. Moreno (Ed.), *The sociometry reader.* Glencoe, Ill.: Free Press, 1960. Pp. 428–436.

Kounin, J., Polansky, N., Biddle, B., Coburn, H., & Fenn, A. Experimental studies of clients' reactions to initial interviews. *Human Relations,* 1956, **9**, 265–293.

Lambert, W. W. Social psychology in relation to general psychology and other behavioral sciences. In S. Koch (Ed.), *Psychology: A study of a science.* (Vol. 6) New York: McGraw-Hill, 1963. Pp. 173–243.

Lewin, K. *Field theory in social science: Selected theoretical papers.* (D. Cartwright Ed.) New York: Harper, 1951.

Lipset, S. M., Trow, M. A., & Coleman, J. S. *Union democracy.* Glencoe, Ill.: Free Press, 1956. Pp. 106–126.

McClintock, C. G., Becker, G. M., & DeLamater, J. D. *Element, aggregates, and levels in social psychological theory.* (SP-243) Santa Barbara, Calif.: TEMPO (General Electric Co.), 1963.

McGrath, J. E. *Integration of small group research information: Program report.* McLean, Va.: Human Sciences Research, 1962. (a)

McGrath, J. E. *A summary of small group research studies.* (HSR-TN-62/3-Gn) McLean, Va.: Human Sciences Research, 1962. (b)

McGrath, J. E. *Systems of information in small group research studies.* (HSR-TN-62/2-Gn) McLean, Va.: Human Sciences Research, 1962 (c)

Mills, T. M. Power relations in three-persor groups. *American Sociological Review,* 1953, **18** 351–357.

Moreno, J. L. (Ed.) *Sociometry and the science of man.* New York: Beacon House, 1956.

Moreno, J. L. Theory of interpersonal networks. In J. L. Moreno (Ed.), *The sociometry reader.* Glencoe, Ill.: Free Press, 1960. Pp. 67–79.

Moreno, J. L. (Ed.) *The sociometry reader.* Glencoe, Ill.: Free Press, 1960.

Mouton, J. S., Bell, R. L., Jr., & Blake, R. R. Role playing skill and sociometric peer status. In J. L. Moreno (Ed.), *The sociometry reader.* Glencoe, Ill., Free Press, 1960. Pp. 388–398.

Nagel, E. *The structure of science: Problems in the logic of scientific explanation.* New York: Har-Court, Brace & World, 1961.

Newcomb, T. M. *The acquaintance process.* New York: Holt, Rinehart, & Winston, 1961.

Northway, M. L., & Detweiler, J. Children's perception of friends and nonfriends. In J. L. Moreno (Ed.), *Sociometry and the science of man.* New York: Beacon House, 1956. Pp. 271–275.

Northway, M. L., & Rooks, M. M. Creativity and sociometric status in children. In J. L. Moreno (Ed.), *Sociometry and the science of man.* New York: Beacon House, 1956. Pp. 194–201.

Pepitone, A., & Kleiner, R. The effects of threat and frustration on group cohesiveness. *Journal of Abnormal and Social Psychology,* 1957, **54**, 192–199.

Pepitone, A., & Reichling, G. Group cohesiveness and the expression of hostility. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory.* (2nd ed.) Evanston: Row, Peterson, 1960. Pp. 141–151.

Raven, B. H., & Reitsema, J. The effects of varied clarity of group goal and group path upon the individual and his relation to the group. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory.* (2nd ed.) Evanston: Row, Peterson, 1960. Pp. 395–413.

Riecken, H. W., & Homans, G. C. Psychological aspects of social structure. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 2. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 786–829.

Schachter, S., Ellertson, N., McBride, D., & Gregory, D. An experimental study of cohesiveness and productivity. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory.* (2nd ed.) Evanston: Row, Peterson, 1960. Pp. 152–156.

Spector, A. J. Expectations, fulfillment and morale. *Journal of Abnormal and Social Psychology,* 1956, **52**, 51–56.

STOTLAND, E., & ZANDER, A. Effects of public and private failure on self-evaluation. *Journal of Abnormal and Social Psychology,* 1958, **56**, 223–229.

TAGIURI, R. Relational analysis: An extension of sociometric method with emphasis upon social perception. *Sociometry,* 1952, **15**, 91–104.

TALLAND, G. A. Task and interaction process: Some characteristics of therapeutic group discussion. In A. P. Hare et al. (Eds.), *Small groups: Studies in social interaction.* New York: Knopf, 1955. Pp. 457–463.

THIBAUT, J. W., & COULES, J. The role of communication in the reduction of interpersonal hostility. *Journal of Abnormal and Social Psychology,* 1952, **47**, 770–777.

THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups.* New York: John Wiley, 1961.

THIBAUT, J. W., & RIECKEN, H. W. Some determinants and consequences of the perception of social causality. *Journal of Personality,* 1955, **24**, 113–133.

ZELENY, L. D. Validity of a sociometric hypothesis: The function of creativity in interpersonal and group relations. In J. L. Moreno (Ed.), *Sociometry and the science of man.* New York: Beacon House, 1956. Pp. 183–193.

(Received February 18, 1965)

Psychological Bulletin
1965, Vol. 64, No. 6, 413–427

ANOTHER LOOK AT PHONETIC SYMBOLISM [1]

INSUP KIM TAYLOR
Lakeshore Psychiatric Hospital,
New Toronto, Ontario

AND

MAURICE M. TAYLOR
Defence Research Medical Laboratories,
Toronto, Ontario

Various possible meanings of "phonetic symbolism" are discussed. Phonetic symbolism is distinguished from onomatopoeia. "Elemental" and "structural" phonetic symbolism are defined. Elemental phonetic symbolism is discussed in terms of 7 hierarchically arranged questions which define "subjective" phonetic symbolism (that detected by observers) and "objective" phonetic symbolism (overrepresentation of particular sounds in words of particular connotations in natural languages). Experimental and empirical evidence relevant to each question are discussed, and it is concluded that both subjective and objective elemental phonetic symbolism are real phenomena, but that the patterns of symbolism are unrelated in historically unrelated languages. The feedback theory of phonetic symbolism (Taylor, 1963) is considered in greater detail than heretofore. The feedback theory asserts that if speakers of a language show subjective phonetic symbolism, then the language will come to show the same pattern of objective phonetic symbolism, and that the pattern of objective phonetic symbolism in the language similarly influences the development of subjective phonetic symbolism in the speakers.

According to the hypothesis of Universal Phonetic Symbolism (UPS), sounds have symbolic connotations which are the same in all languages. Taylor (1963), rejecting this hypothesis as no longer tenable, contended that phonetic symbolism was strong in probably all languages, but that the actual connotations of sounds differed among unrelated languages. She further proposed a feedback theory of language-dependent phonetic symbolism.

Weiss (1964) has criticized both Taylor's grounds for rejecting the UPS hypothesis and her theory. Though Weiss' specific criticisms may be readily answered, they exemplify some common confusions among the different meanings of the term phonetic symbolism. This paper attempts a clarification and classi- ication of these usages. In the process, it examines anew the evidence for different orms of phonetic symbolism, and presents he feedback theory in extended form.

WHAT IS PHONETIC SYMBOLISM?

Phonetic Symbolism and Onomatopoeia

Several studies, following Sapir (1929), have required subjects to judge nonsense syllables as if they referred to various attributes, such as size, brightness, pleasantness, and so forth. The best known result of these experiments is that for English speakers the vowel influences the "feel" of a nonsense syllable in respect to size. High vowels, such as /ɪ/, tend to give a feel of smallness, while lower vowels tend to give a feel of greater size. The favorite rationale for the effect of vowels on the size feel of a nonsense syllable has been that small objects tend to make higher pitched sounds than do larger objects. High-pitched vowels, therefore, seem appropriate for small objects. The sound symbolism is, according to this rationale, a subtle form of onomatopoeia.

Onomatopoeia is defined in dictionaries as the generation of words in imitation of the sounds made by the thing or action designated. The relationship in the literature between phonetic symbolism and onomatopoeia can be at best described as confusing. At one extreme, phonetic symbolism was discussed under the heading onomatopoeia without ever being labeled as phonetic symbolism by Osgood (1965), as though phonetic symbolism *is* onomatopoeia. At the other extreme, Brown (1958) clearly distinguished the two phenomena:

> We all know about onomatopoeia—the imitation of non-speech sound in speech. But this is not onoma-

---

[1] The opinions expressed in this paper are those of the authors and do not necessarily represent the views of the Canadian Department of National Defence.

> topoeia. Here speech sounds suggest spatial and visual dimensions, size and brightness [p. 114].

Normally speaking, one of the more striking characteristics of natural sounds is rhythm or lack of rhythm. Hence, onomatopoetic words tend to be rhythmic or not, depending on the thing designated. It is, a priori, reasonable to suppose that the nonrhythmic symbolism that connects size with vowel sounds is also a form of onomatopoeia. However, it is not so reasonable to suppose that onomatopoeia generates the affinities that have been found between consonants and size, or between vowels and, say, pleasantness. Neither is it reasonable to suppose that the same onomatopoetic effect makes Tamil and Japanese speakers feel /ɪ/ to be appropriate to large objects and /U/ to small, unless in these countries large, instead of small, objects emit high-pitched sounds.

> Then, there are the results of Eberhardt's (1940) experiments with deaf children to be accounted for. Eberhardt, using the Sapir-Newman procedure with English reading deaf children obtained similar orderings of English vowels as for hearing (English-speaking) subjects. The deaf subjects obviously could not have based their judgment on the tonal quality of the sounds [Taylor, 1963, p. 206].

There is obviously a sound symbolism that is not related to onomatopoeia. This is the effect called "phonetic symbolism."

Elemental and Structural Phonetic Symbolism

Thus far, we have considered both phonetic symbolism and onomatopoeia as if their effects depended only on the sounds in the words or syllables, without consideration of the order or relationship of the sounds. In Sapir's material the structural pattern C-C (initial consonant and final consonant) of "*m-l*" remained constant while one sound element, the middle vowel, was varied, so that we can safely attribute the varying size connotations of the syllables to the varying sound element.

It is difficult to determine whether or not elemental onomatopoeia exists. The words that seem particularly appropriate for description of things with clear acoustic concomitants do not necessarily contain sounds like those they describe. Consider the low vowels in "cock-a-doodle-doo" and compare them to the high screech that the word is supposed to imitate. In other languages, that same screech is expressed with entirely different sounds.

Experimentally, one cannot demonstrate onomatopoeia of any kind, because of the ever-present possibility of phonetic symbolism. By chance, or for cause, the sounds appropriate by reason of phonetic symbolism may be like the sounds described. Casual analysis would suggest them to be appropriate by reason of onomatopoeia. One can only illustrate onomatopoeia, never demonstrate it.

While experimental demonstration of elemental (or even structural) onomatopoeia seems difficult, experimental demonstration of structural phonetic symbolism appears possible in principle. Structural phonetic symbolism may be distinguised from structural onomatopoeia by the same tests that apply to elemental phonetic symbolism. Can structural symbolism be demonstrated for attributes with no apparent acoustic referents? For a single attribute, do speakers of unrelated languages agree on the feel due to structure? So far as we are aware, no studies have been directly aimed at these questions.

We shall enquire in this paper only into questions of elemental phonetic symbolism, though the other questions will, from time to time, be raised where they intrude on the main topic of enquiry.

Unit of Phonetic Symbolism

The effects of elemental phonetic symbolism are nominally due to the sounds in words. However, as has often been pointed out,

> . . . the total number of possible sounds is greatly in excess of those actually in use. Indeed, an experienced phonetician should have no difficulty in inventing sounds that are unknown to objective investigation [Sapir, 1921, pp. 44–45].

Furthermore, the sounds that are heard are far fewer in number than the sounds that are produced:

> Under normal circumstances, in the decoding process, people do not distinguish difference between allophones . . . . Consequently the allophone is too small to be a unit in the encoding or decoding process, implying that the phoneme is [Saporta, 1965, p. 62].

No only do listeners fail to hear fairly large differences in spoken sounds, they also hear nonexistent differences between two identical sounds in different contexts (Broadbent & Ladefoged, 1960; Ladefoged & Broadbent, 1957).

All these failures of hearing imply that the natural unit for analysis of elemental phonetic symbolism is the phoneme. All studies using acoustic stimulation, save possibly that of Miron (1960, 1961), have indiscriminately allowed allophonic variation within the presented phonemes. Similarly, studies using written materials have of necessity used phonemes as the units of analysis. In a sense, it would be better to talk of "phonemic" symbolism, to allay the suspicion that we might actually be enquiring into the symbolism of speech sounds as such.

Problems in Comparing Phonetic Symbolism across Languages ·

Sounds that belong to a single phoneme in one language may be divided among two or more phonemes in another. Sounds that are phonemically valid in one language may not even occur in another. It is impossible to map the phonemes of one language in a uniquely meaningful way onto those of another. Though phonemes do not agree in detail across languages, the main features of some phonemes match well enough across languages that they may be meaningfully treated for some purposes as the same phoneme. The English /t/ and the Korean /ㅌ/ do not include the same roster of sounds, but each is an unvoiced apico-dental stopped consonant. Most allophones of /ㅌ/ would sound like acceptable variants of /t/ to the ear of a normal English-speaking person untrained in phonetics. In this sense, most language pairs include many matched phonemes.

Since the phoneme library of one language does not match that of another language, it is meaningless to enquire about completely universal phonetic symbolism. However, it is meaningful to consider whether phonetic symbolism gives to matched phonemes the same feel in both languages. It is also meaningful to ask questions in which the unit of analysis is more gross than the phoneme. For example, the results of Sapir (1929), quoted above, suggest the generalization that low vowels give large feel to nonsense syllables. Such a generalization gives rise to a valid question in any language that uses vowels, without implying a vowel to vowel correspondence among languages. Phonetic symbolism could in principle be universal in terms of generalizations of this type, even though direct phoneme matches are not possible over all languages.

Phonetic Symbolism Felt by Observers, and Phonetic Symbolism Existing in Natural Languages

One question of phonetic symbolism is whether some sounds feel appropriate to certain concepts even in nonsense syllables. Sapir's finding that /ɪ/ in MIL seems appropriate to the concept of "smallness" supplies a partial answer. A separate question is: "Are some sounds found proportionately more in words of a certain meaning than in words of other meanings?" For example, Jespersen (1922) noted the frequent occurrence of the letter *i* in English words with the connotation of "small," and of *u* in words with the connotation of "contemptuous." These two questions can be asked, and answered, independently of each other, although even recently Miron (1960) and Weiss (1964) discuss these two versions of phonetic symbolism as if they were logically dependent. But there is no a priori reason why the English language should use *i* more among small than large words, even if English speakers find that /ɪ/ gives nonsense words a small feel.

If both subjective phonetic symbolism (symbolism felt by observers) and objective phonetic symbolism (based on the distribution of sounds in a language) are demonstrated, then we can ask: "Is the sound felt to be appropriate to certain concepts in nonsense syllables also found in real words of the same concept more often than chance would indicate?" For example, /ɪ/ is shown by Sapir to be appropriate to "smallness" in nonsense syllables and is noted by Jepersen to occur frequently in "small" English words.

TABLE 1

SPECIFIC EXAMPLES OF SEVEN QUESTIONS OF ELEMENTAL PHONETIC SYMBOLISM

A. Association felt by individuals between individual sounds and their symbolic connotations in arbitrary contexts (subjective phonetic symbolism).	
1. An individual	Does John Smith always feel that /ɪ/ in a nonsense syllable, as in MIK, makes the syllable sound small?
2. Speakers of one language community	Do individuals speaking the same language agree on (1)? Do all English-speaking people agree that /ɪ/ in MIK makes it sound small?
3. Speakers of unrelated languages	Do individuals speaking unrelated languages agree on (1)? Do speakers of Japanese, Korean, Tamil, etc., all agree that /ɪ/ in MIK makes it sound small?
B. Frequent occurrence of individual sounds in words of certain concepts in natural languages (objective phonetic symbolism).	
4. Within a language	In English, is /ʌ/ found more often than chance in "unpleasant" words such as *Grumble?*
5. Across unrelated languages	Is the same /ʌ/ found too often in unpleasant words in Japanese, Korean, Tamil, etc.?
C. Correspondence between (A) and (B), or subjective and objective phonetic symbolism.	
6. Within a language	Is /ɪ/, which is felt by English speakers to make MIK sound small, also found too often in "small" words in English?
7. Across unrelated languages	Is /ɪ/, which is felt by English speakers to make MIK sound small, also found too often in "small" words in Japanese, Korean, Tamil, etc.?

Seven Questions of Phonetic Symbolism

When subjective phonetic symbolism and objective (language content) phonetic symbolism are considered as independent subjects of enquiry, there are seven questions that may be asked about the generality of phonetic symbolism:

1. Does a particular individual find a consistent feel in words or nonsense syllables containing certain phonemes when these words are judged in respect of certain attributes? (Can subjective phonetic symbolism be demonstrated within an individual?)

2. Can consistent subjective phonetic symbolism be demonstrated across individuals of a single language community?

3. Can consistent subjective phonetic symbolism be demonstrated across individuals speaking unrelated languages?

4. In a single language, do particular phonemes occur more often in words associated with certain concepts than in the words of the language as a whole? (Does a particular language show objective phonetic symbolism?)

5. Do unrelated languages display the same objective phonetic symbolism?

6. Does the subjective phonetic symbolism displayed by speakers of one language have the same pattern as the objective phonetic symbolism found in that language?

7. Does the subjective phonetic symbolism displayed by speakers of one language have the same pattern as the objective phonetic symbolism found in an unrelated language?

The import of these general questions may be clarified by the specific examples displayed in Table 1.

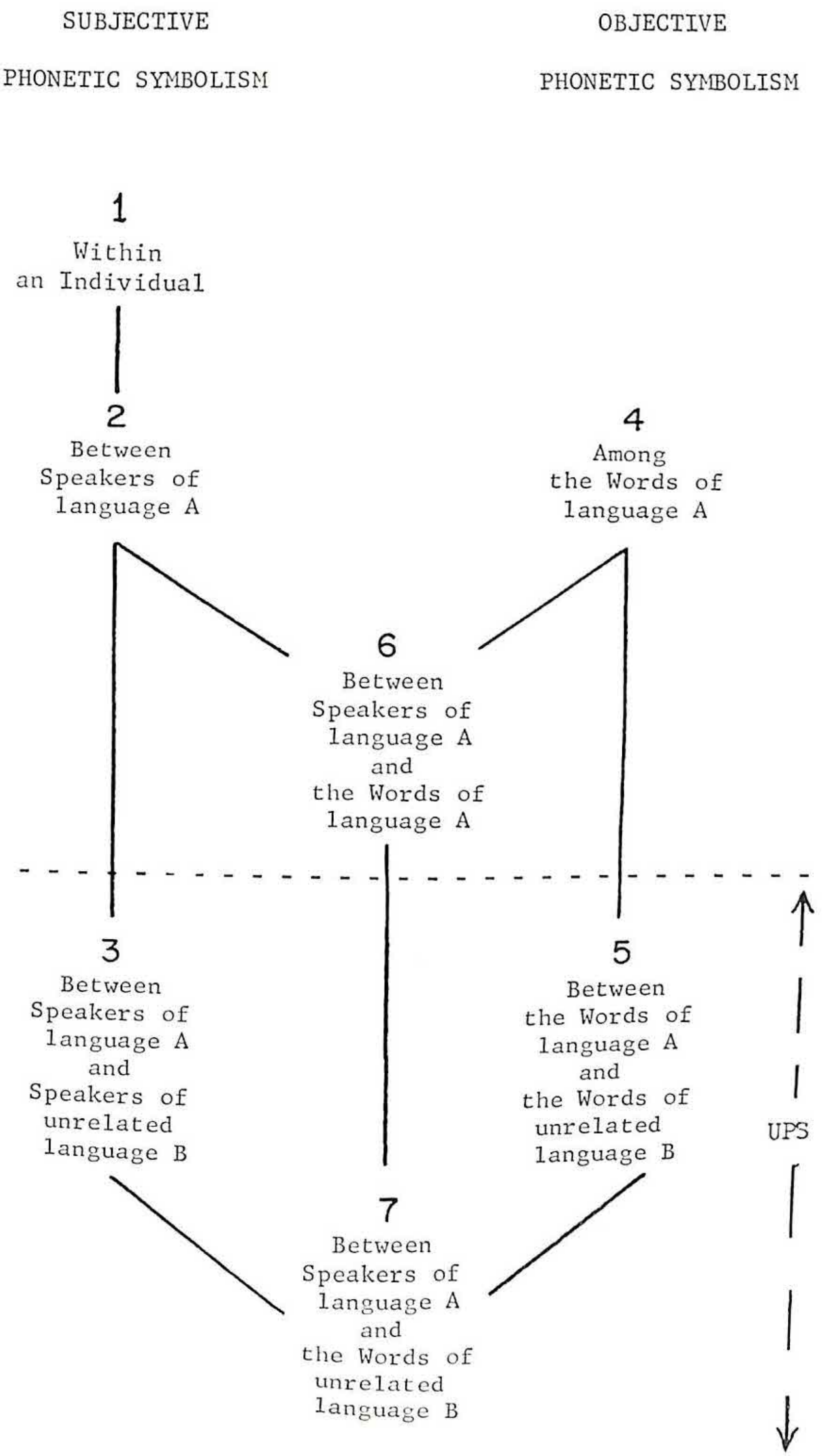

SUBJECTIVE
PHONETIC SYMBOLISM
OBJECTIVE
PHONETIC SYMBOLISM
1
Within an Individual
2
Between Speakers of language A
4
Among the Words of language A
6
Between Speakers of language A and the Words of language A
3
Between Speakers of language A and Speakers of unrelated language B
5
Between the Words of language A and the Words of unrelated language B
7
Between Speakers of language A and the Words of unrelated language B
UPS

FIG. 1. Logical relationships among the seven questions of elemental phonetic symbolism. (If two questions are connected by a line, a yes answer to the lower question implies a yes answer to the higher, and a no answer to the higher implies a no answer to the lower.)

The questions form a hierarchic tree as in Figure 1. Three questions (1–3) deal with successively more general questions of subjective phonetic symbolism. Answers to any of these questions imply answers to others. For example, if subjective phonetic symbolism could not be demonstrated in any individual, then it would be meaningless to ask whether it had the same pattern for individuals speaking unrelated languages. Two questions (4 and 5) deal with objective phonetic symbolism, and they are related to each other in the same way as Questions 2 and 3. Two questions (6 and 7) connect subjective and objective phonetic symbolism. Question 7 is meaningless if any one of Questions 3, 5, and 6 is answered negatively. Similarly, a positive answer to Question 7 implies a positive answer to all the other six questions.

The seven questions, together with the phonetic symbolism-onomatopoeia and the elemental-structural dichotomies, serve to define some of the meanings of what we take to be phonetic symbolism.

We shall discuss the evidence regarding elemental phonetic symbolism in terms of each question separately.

EVIDENCE ON ELEMENTAL PHONETIC SYMBOLISM

Much of the evidence has been reviewed by Taylor (1963), but without the present classification scheme. A new approach to some of the evidence and inclusion of newly published evidence are required both by the new classification scheme, and by the appearance of Weiss' (1964) criticisms of the earlier review. We shall consider evidence which we or other authors suppose to be relevant to the seven questions of elemental phonetic symbolism. Evidence is considered in respect of each question in turn, though because of the logical relationships among the questions, clear answers to some questions provide evidence for the discussion of other questions.

Subjective Phonetic Symbolism

QUESTION 1: *Can subjective phonetic symbolism be demonstrated for a single individual?*

Question 1 has not been deliberately studied. However, the very clear positive answer to Question 2 confirms that individuals do demonstrate subjective phonetic symbolism.

QUESTION 2: *Do individuals speaking the same language show the same pattern of subjective phonetic symbolism?*

This question has been the subject of several studies (Dagiri, 1958; Miron, 1960, 1961; Newman, 1933; Sapir, 1929; Taylor & Taylor, 1962), all of which used the analytic method to determine the connotations of individual phonemes, and each of which produced a decisive positive result. All the studies included the size attribute, Newman added brightness, Taylor and Taylor included movement, warmth, and pleasantness, while Miron used 15 different attributes. All studies except Dagiri's included English speakers, while Dagiri, Miron, and Taylor and Taylor used Japanese speakers with different degrees of English knowledge. Taylor and Taylor also used Tamil and Korean speakers. In the five experiments quoted, there have been 51 independent tests of the consistency on different conceptual scales of subjective phonetic symbolism across subjects. All have shown positive correlations across speakers speaking the same language, and only two, or possibly three, conceptual scales have failed to give strong correlations (Miron, English speakers, "heavy-light" and "cold-hot"; Taylor and Taylor, Tamil speakers, "movement"). This overwhelming support for the effectiveness of phonetic symbolism suggests that the symbolism is a pervasive feature of language behavior. It is not restricted to this or that attribute, or to trained subjects, or to speakers of this or that language. Wherever we look, we find elemental subjective phonetic symbolism.[2]

QUESTION 3: *Do individuals speaking unrelated languages agree on the connotations of a given sound?*

---

[2] After this manuscript was completed, S. Ertel (1965) kindly sent a prepublication proof of his article in which he used a technique like the analytic method described for Question 2; except that arbitrary sound complexes formed the poles of judgment scales, with meaningful words to be placed on the scales. The results again support the "yes" answer to Question 2.

Question 3 has been attacked directly by Miron (1960, 1961), Davis (1961), and Taylor and Taylor (1962). Weiss (1964) presents the Miron and Davis experiments as contradicting Taylor and Taylor's finding that speakers of different languages do not agree. We do not agree that there is any contradiction among the results of the three studies.

1. Davis asked English- and Swahili-speaking children to match the words TAKETE and ULOOMU, nonsense in each language, to two drawings. Both groups of children agreed that TAKETE went with the angular drawing.

For several reasons, this result can carry very little weight in a discussion of phonetic symbolism. First, only one word pair was used, so that the chance of agreement between groups was .5 a priori, if each group showed internal agreement. Secondly, TAKETE and ULOOMU differ not only in their phonetic elements, but also in their structure (CVCVCV versus VCVCV); and hence in their rhythmic effects. The results may be valid, but one cannot conclude unequivocally that the elemental phonetic factor was responsible for the results. Thirdly, the words are almost onomatopoetic (probably this is why the two words were selected): TAKETE is similar to the sound made by a small angular object bouncing around, while ULOOMU is less similar, but surely suggestive of a soft or liquid mass moving slowly.

2. Miron (1960, 1961) and Taylor and Taylor (1962) used very similar techniques. Miron asked English and Japanese speakers to listen to spoken CVC syllables, mostly nonsense, and to judge the syllables on 15 scales of meaning. He found an overall average correlation of .55 between the English and the Japanese ratings. This correlation was highly significant, enabling him to conclude that there was considerable communality of symbolism between the groups.

Like Miron, Taylor and Taylor presented CVC syllables, mostly nonsense, to speakers of English, Japanese, Korean, and Tamil, and required them to judge the syllables on four scales of meaning. Their syllables were written with the appropriate alphabets or syllabaries. They computed the symbolic value on each conceptual scale of each consonant in the initial and final positions separately, and of the vowels, and correlated the values across pairs of languages. They found correlations of about .24 between English and each other languages ($p < .05$). Had they, like Miron, used only English and Japanese they also would have concluded that there was a common factor in phonetic symbolism. However, they found no significant correlation between pairs of languages not including English. They attributed the positive correlations between English and each other language to the fact that subjects were not truly monolingual, having been exposed in various ways to English (United States military and civilian personnel frequent Tokyo and Seoul to such an extent that some shop signs are in English in each city; English is an official language in India, and English is a school subject in all three countries.) Where there was no such direct exposure, between Tamil and Korean, Japanese and Tamil, or Japanese and Korean, the correlations were essentially zero. Thus, unlike Miron, Taylor and Taylor concluded that there was no commonality of symbolism among the languages, and that the strong phonetic symbolism displayed by each group must be language-dependent.

The detectable effect of minor contamination by English exposure of Japanese children in Tokyo must raise suspicions about Miron's higher correlations between Japanese and English scores. Miron used Japanese academics who had, on the average, been in the United States more than a year and hence had been exposed far more strongly to English than had Taylor and Taylor's subjects. The two studies considered together do not contradict one another. They do suggest that further studies of subjective phonetic symbolism must control the degree to which subjects have been exposed to languages other than their own, especially to other test languages.

The answer to Question 3 seems to be "no."

Auditory versus Visual Presentation of Test Material

The question of presentation method is pertinent for Question 7, but arises for the first time under Question 3. It will be considered here only, for both types of study.

1. Auditorily presented real words used in word-matching experiments are usually pronounced by native speakers and tape-recorded. Real words have meanings, and these meanings may influence the native speakers who pronounce the words. According to Eberhardt (1940), who used deaf subjects, if the deaf were unaware of the meanings of TAP and POUND, they spoke the two words with little difference, but once they learned the meanings, TAP became lighter and softer, while POUND took on a "richer resonance." In addition to the meanings of the words, the speaker's own susceptibility to the symbolic connotations of words or nonsense syllables may influence his way of pronouncing the words. Jespersen (1922) observed that we have a tendency to lengthen and strengthen single sounds under the influence of strong feeling or in order to intensify the effect of the spoken word; thus, in "extremely LONG" the vowel O may be lengthened. As a speaker is not usually aware of these influences or of the purpose of the experiment (as is usually the case in word-matching experiments), he would not be expected to eliminate them.

There have been two experiments in which different methods of presentation were directly compared. Both were word-matching studies (see Question 7, below) and the results are therefore to be interpreted with caution. Brown, Black, and Horowitz (1955) compared two methods: visual alone and audio-visual combined. They concluded that the two methods produced similar results, with correlations on individual items in three different languages ranging from .55 to .61 ($p < .02$). If anything, the visual-alone group showed an overall performance slightly more successful than the audio-visual combined group. Brackbill and Little (1957) compared three methods: auditory, visual and audio-visual. The three methods differed slightly ($p < .05$) in producing accurate matching, with the audio-visual and visual groups being superior to the audio group. In the English-foreign languages comparison, the visual group showed a tendency to a higher percentage of correct guesses; but in the foreign-foreign language matches, the opposite seemed to hold.

Ertel and Dorst (1956) used only auditory presentation, and suggested in discussion of their positive results that expressive voice quality of speakers could have influenced the subjects' responses.

2. Auditorily presented nonsense syllables are not always totally nonsense. They can be assigned association values and are meaningful in various degrees. Their associations may influence the pronouncer of the syllables. When the pronouncer is also the investigator as were Sapir (1929) and Newman (1933), the investigator's own hypothesis on phonetic symbolism also may influence the expressive quality of pronunciation. Newman showed that the symbolic value of accenting the vowel is heavily in favor of largeness for English speakers.

When several languages are involved, it becomes almost impossible to pronounce sounds that will be equally valid phonemically in all the languages. The pronouncer can entirely disregard phonemic systems of the languages and produce some arbitrary sounds, but these sounds may turn out to be closer to phonemes in some languages than to others. In Miron's (1960, 1961) study, only two languages were involved. Even so, a trained phonetician had a delicate and difficult task of attempting a compromise in producing sounds so that the pronounced sounds would not favor either of the languages, Japanese or English. What this attempt at a compromise achieved is hard to know.

3. Visually presented real words of foreign languages are usually spelled in Romanized alphabets for the English-speaking subjects. Some languages have sound systems that can be more closely approximated than others in such a Romanization attempt.

One of the most conspicuous features of real words is visual length. An indirect attempt has been made by Brown, Black, and Horowitz (1955) to determine if the visual length could have influenced the matchings. By examining the data, they decided that subjects did not rely on visual length to match foreign words to English words. On the other hand, Brackbill and Little (1957); after examining their data for the factor of visual length and other factors such as vowels, consonant types, spacing of compound words, and connotation; found that all these factors markedly influenced the agreement among the subjects. As a consequence, Brown et al. concluded that universal phonetic symbolism existed, while Brackbill and Little concluded contrary to the hypothesis.

4. Visually presented nonsense syllables have not, as yet, been widely used. So far apparently only Taylor and Taylor (1962) have employed written stimuli in an analytical study. For their experiment, Taylor and Taylor prepared the written stimulus syllables in four different languages using the alphabets and syllabaries of the languages.

One of the dimensions on which the syllables were judged was size, and visual length was investigated in one of the preliminary experiments with English-speaking subjects. Syllables written in a longer form were judged as having "bigger" connotations. However, visual length is easy to control, and in the main experiment it was equalized in all the languages.

Even after visual length has been taken care of, there remains the differences in visual forms of different letters in one language and of phonemes under comparison in different languages. Could these differences in visual forms of phonemes have influenced subjects' judgments of symbolic connotations of phonemes? One can never unconfound the effects due to pure phonemes and possible effects due to these visual forms of phonemes, as long as the material is visually presented.

When Taylor and Taylor's graphically presented study is compared with the auditorily presented studies of Miron, Newman, and Sapir for dimensions of meanings and sounds in common, the results seem to be in rather good agreement. This observation suggests that perhaps the two methods of presentation do not differ too markedly.

It is hardly worthwhile to generalize from the results of comparisons involving real words to comparisons involving nonsense syllables. For one thing, the results from real words are conflicting; and secondly, the condition of comparison is somewhat different in the studies involving two types of test material. Unlike nonsense syllables which can be altered in a controlled way, the numerous features of real words cannot be controlled before a study but must be examined, often haphazardly, after the study is completed.

Objective Phonetic Symbolism

QUESTION 4: *Does a particular language show objective phonetic symbolism?*

Jespersen (1922) pointed out that in many Indo-European languages /ɪ/ is used in a high proportion of the words connoting smallness, and /ʌ/ (as in BUT in English) in words with an undesirable or contemptuous connotation. This is an intuitive affirmative answer to Question 4 in respect to English and other Indo-European languages. Householder (1946) concurred in respect to /ʌ/, but his technique has been called "dubious" (Miron, 1960). Marchand (1958, 1959) has extended to over a hundred the list of concept-tied phonemes and phoneme clusters in English, but the technique remains subjective. With presently available computing equipment, and the developing apparatus of machine translation, it might be possible to give a more satisfactory direct answer to Question 4, but for the present we must be satisfied with a tentative "yes."

QUESTION 5: *Do unrelated languages display the same objective phonetic symbolism?*

If Questions 2 and 4 are answered affirmatively, yes answers are in principle possible to Questions 5 and 6. No evidence has been produced relevant to Question 5 because languages unrelated to English have not been examined in the same detail as have the Indo-European languages.

Correspondence between Subjective and Objective Phonetic Symbolism

QUESTION 6: *Is the subjective phonetic symbolism of a language community reflected in words of that language?*

Question 6 has been considered by Newman (1933), Berko and Brown (cited by Brown, 1958), and Taylor (1963). Newman produced a list of words having connotations of smallness and of largeness, and examined all the sounds of which they were composed in the light of his analytical determinations of their symbolic size values. He concluded that the subjective phonetic symbolism of his observers was not reflected in the words of English. Berko and Brown considered the frequency of usage of words, and agreed with Newman.

Taylor used the result of Taylor and Taylor (1962) that the symbolic value of a sound depends to some extent on its position in the word, and used their values for the size symbolism of the initial consonants in testing Newman's list. She concluded that "small" consonants were more likely than "large" consonants to start "small" words, and vice versa for large words. Weiss criticized this finding on the grounds that the words in the list might have been specially selected. The criticism is not valid. To make the list, Newman found all the words in Roget's Thesaurus under the headings "Greatness, Smallness, Size, and Littleness" and struck out "the repetitions, derivatives and phrases." Eleven judges were asked to assess the 500 words remaining, and a majority decision was taken as to whether each word belonged in either list and, if so, in which list it belonged. Selection on a phonetic basis seems unlikely.

Taylor originally considered for illustrative purposes only the initial consonants in Newman's list. Since then, we have considered the first vowel in Newman's initial consonant words, using the vowel scores from Taylor and Taylor's CVC syllables. According to these scores, /ɪ/ and /ɛ/ are the small vowels, /u/ and /o/ the large. Of the words in Newman's list having small vowels as their first, 78 are "small" and 37 "large." Of the words having large vowels as their first, 9 are "small" and 18 are "large." The difference in ratio is even more marked than that for the initial consonants, and is highly significant ($p < .002$). This result again suggests an affirmative answer to Question 6, as far as the English language is concerned, and therefore supports the intuitive affirmative answer given by Jespersen (1922), Householder (1946), and Marchand (1958, 1959) to Question 4.

QUESTION 7: *Does the subjective phonetic symbolism displayed by speakers of one language have the same pattern as the objective phonetic symbolism found in an unrelated language?*

Question 7 is in a sense the "grand question" of UPS. By implication it asks whether all observers and all languages share common connotations for given sounds. In relation with other questions, positive answers to Questions 1, 2, 3, 4, 5, and 6 are necessary, and sufficient for a yes answer to Question 7. However, Question 5 has no answer in the literature to date, and we have concluded that a no answer must be given to Question 3. Question 7 has been directly studied using variants of word-matching technique, mostly using English-speaking subjects.

Word-Matching Experiments

In a standard form of word-matching experiments, speakers of a given language, say English, are asked to choose the correct translations into their own language of the foreign word pairs presented. In a variant, Ertel and Dorst (1965) asked German subjects to designate which member of a foreign word pair belonged to one of the two opposite concepts that were defined by a set of words in German. The same set of words defining concepts were translated into foreign words to be judged.

Most of Weiss' criticisms are directed at Taylor's interpretation of the standard word-matching studies. After reviewing the then available word-matching studies, some with positive and some with negative results, Taylor noted that the spurious positive results were probably due to procedural anomalies. Weiss takes issue with this interpretation, suggesting that perhaps negative results were due to inappropriate experimental conditions.

We recapitulate and expand here Taylor's objections to the word-matching technique:

1. Translation bias: Taylor's main objection to word-matching as a technique applies to all word-matching studies currently in the literature, as long as they involve real words in foreign languages: somebody has to decide what words are appropriate translations of what. If the translator is involved with the experiment, he may consciously or unconsciously bias his selections. Rosenthal (e.g., 1963) has documented many empirical

and experimental cases in which such experimenter bias influenced the results even in fairly strictly controlled conditions. The Tsuru and Fries (1933) experiment is subject to this criticism, Tsuru having translated the test words himself, as is that of Maltzman, Morrisett, and Brooks (1956) who used Tsuru's list.

An obvious but naive control against this bias is to use a translator who is not aware of the purpose of the experiment. This control has been adopted in most of the word-matching studies since Allport (1935) introduced the service of a translator for preparing stimulus material. Unfortunately, this "control" increases the likelihood of a bias in favor of correct matching.

> There seldom exists one and only one foreign equivalent to an English test word. In the process of translating English test words, a translator . . . could choose the foreign equivalent that more resembles or clang-associates with the given English word. For example, in translating OLD into a Japanese equivalent, a translator may choose OITA instead of TOSHITOTTA because O in OLD makes him clang-associate to a Japanese word that also has a beginning O [Taylor, 1963, p. 202].

Besides, OITA is nearer in length to OLD than TOSHITOTTA, which may also bias the selection in favor of OITA. Tsuru and Fries' list did in fact use OITA.

2. Correct matching does not allow an unequivocal interpretation that UPS is responsible. This second criticism applies to all forms of matching studies, including Davis', which involved nonsense words but still used the matching technique. Weiss (1964) admits as much, and Brackbill and Little (1957) even concluded that phonetic symbolism was not responsible for correct matching. As late as 1960, McMurray could suggest only meager bases—possibly vowels on one dimension, and hard consonants on a second of his three dimensions of meaning—for his positive results. Ertel and Dorst (1965) merely attributed correct matching to expressive sound symbolism, without discussing *what* "sound complexes fit in sound character to [which] respective word-matching."

Now if correct matching in an unbiased set of translations could be demonstrated, it is quite likely that the contrasts on which the subjects based their choice might be contrasts of structure, length, or, in auditory presentation, expressive voice quality, or some other factors. There are probably linguistic laws analogous to Zipf's (1935) law that common words (most of the studies for some reasons avoided uncommon words) tend to be short. It is not unlikely that words similarly used might be similar in structure in different languages purely from the operation of laws concerning the economics of information transfer. It was this idea, expressed differently, that led Tsuru and Fries (1933) to originate the word-matching technique. They wished to demonstrate the Gestalt position that "meanings are properties of naturally organized wholes or processes."

3. Difficulty of associating specific sound with specific symbolic connotation: Supposing Objections 1 and 2 were satisfactorily taken care of, there would still be the problem of associating specific sounds with specific symbolic connotations. One of the central issues of UPS is to determine how certain sounds are appropriate to certain meanings. To take an example from Weiss (1963), the FAST-SLOW pair was correctly matched to K'UAI-MĀN (Chinese) and TEZ-SUST (Hindi). Here, FAST, K'UAI, and TEZ or their antonyms SLOW, MĀN, and SUST do not share any comparable phonemes, either for consonants or vowels at any position of the words. Sometimes phoneme matches are even conflicting—/a:/ is used for FAST in English, and /a/ in MĀN for slow in Chinese. If some of the phonemes there actually do carry the fast or slow connotations in the manner of /ɪ/ carrying small connotation, it is hard to pinpoint which sound is carrying which connotation. The difficulty is compounded if the extension to other languages is considered—VITE-LENT (French), PRESTO-LARGO (Italian), PALLEUN-NURIN (Korean), HAYAI-OSOI (Japanese), SCHNELL-LANGSAM (German), SKORII-MEDLENNI (Russian), RAPIDO-LENTO (Spanish).

These three problems make the design of a valid word-matching test of UPS difficult. They have not been solved in the studies reported to date.

4. Method of reporting: A fourth problem of the word-matching experiments now in the literature concerns the method of reporting

the results. Generally, the measure reported is the overall percentage of correctly matched pairs. This measure is a compound of the amount of agreement between subjects and the number of pairs matched more often correctly than not. For example, 75% correct matching could be produced if subjects always agreed, and got 75% of the pairs always correct, or if each pair were correctly guessed by 75% of the subjects. These two extreme possibilities would lead to quite different interpretations of the data. As the percentage of correct guessing usually reported is 50%–60%, this question assumes major importance.

We know from many studies (see under Question 2, above) that English-speaking observers will agree fairly well on most of the pairs, so the correctness of their agreement is the only issue in question. Since most of the lists used about 20–30 word pairs, at least 70% of the pairs must be matched more often correctly than not for the pair-matching result to be significant at the .05 level for a single list. No study since Tsuru and Fries (1933) has reported a list matched better than 65% correct overall, and only 3 of 21 lists have given over 60%. Brown (1958) reports that in the Brown, Black and Horowitz (1955) study, the conceptions were correct twice as often as they were incorrect. Those authors analyzed Allport's (1935) data in similar fashion and found comparable results. They are thus claiming 66% accuracy. Weiss (1963) with four 20-pair lists selected from Brown, Black and Horowitz's study, obtained 62% matched more often correctly than not, with about 55% correctly matched overall. In 80 essentially independent pairs, he found 50 correctly matched more often than not. This is barely significant at the .05 level. These results suggest that some level around 65% probably represents the probability of a pair being matched more often correctly than not under typical conditions. In other words, if a biased translation happened only once in 15 words ($7\frac{1}{2}$ pairs), the deviation from chance would be entirely explained.

The results of current word-matching studies cannot be considered as giving evidence in favor of agreement of connotation between English and the various foreign languages used. Brown (1958) has mentioned a minor study that further discredits the hypothesis. When he presented the word list used by Brown, Black, and Horowitz to Chinese-speaking subjects in Boston, he obtained slightly (not significantly) better than chance matching overall, but the Chinese subjects disagreed significantly with the English subjects in their direction of agreement on the different pairs. Brown says "the results as a whole do not strengthen our faith in a universal human [phonetic] symbolism."

All in all, the results of experiments on Question 7 using word-matching type technique may be summarized as having involved much effort, perhaps in the wrong direction, and produced little evidence directly relevant to UPS. Such as it is, the evidence points to a negative answer, as should be anticipated from the negative answer to Question 3.

TAYLOR'S FEEDBACK THEORY OF PHONETIC SYMBOLISM

As a background to discussion of the development of phonetic symbolism, we shall postulate an extreme case. Suppose a language is produced without reference to existing languages, and is used as the sole language of some community. The vocabulary of this language is selected by randomly assigning sound combinations to concepts. We wish to suggest that, over time, even this language will develop phonetic symbolism and that the speakers will show subjective phonetic symbolism when tested with nonsense syllables. By extension from this extremely unfavorable case, we postulate that the same mechanisms are responsible for the varieties of phonetic symbolism actually found.

In our hypothetical protolanguage, sounds occur at random in association with concepts. Sampling fluctuations ensure that some sounds will actually occur more often in connection with particular concepts than in the total vocabulary of the language. For example, let us suppose that the language has 20 words with the connotation of hot, and that / R / is the initial letter in five of them. If / R / is the initial letter of less than 25% of the words of the language, then it is overrepresented as the initial letter of "hot" words.

We have ascribed the overrepresentation of sounds in words of some concepts solely to sampling fluctuations. In natural languages, of course, discrepancies between actual and chance distribution are much greater. The fact that many words associated with one concept derive from the same historical root ensures the overrepresentation of the sounds in the root word in the daughter words associated with the root concept. For example, consider in English the persistence of / v / in words dealing with words: verb, verbal, verbose, verbiage, and so forth. The very existence of language families is discovered because of the existence of such patterns.

If, in our hypothetical protolanguage, a person uses a word with the connotation of heat, he is likely to use a word starting with /R/. Whenever this happens, /R/ and "heat" tend to become associated with one another. If /R/ is also overrepresented in "cold" words, then /R/ and cold also become associated. There would then be no bias toward /R/ feeling cold or warm, but it would be associated with temperature in general. However, if /R/ happened to be underrepresented as the initial letter of cold words, then its temperature-related associations would be to heat. The speaker of the language would feel /R/ to be proper in a "hot" word, and might feel that it was in a way wrong in a cold word. He has developed subjective phonetic symbolism for /R/ on the hot-cold scale of meaning.

If the speaker has developed such subjective phonetic symbolism, it will influence him in the choice of words when he speaks. To give the impression of heat, he can use /R/ words, even if they do not directly denote temperature. When he wants to use a word directly related to heat, /R/ words will seem appropriate. Hence the effective vocabulary becomes biased so that the proportion of /R/ words used for heat increases, and the proportion used for cold decreases. This increase in bias aids the development of subjective phonetic symbolism in other speakers, who are brought up using the changed language.

The thesis is that if objective phonetic symbolism exists in a language, then the speakers will develop subjective phonetic symbolism: If speakers have subjective phonetic symbolism, then the language will develop objective phonetic symbolism. The two patterns will be identical.

This direct analysis of the effects of feedback on phonetic symbolism suggests that each sound should eventually become maximally associated with one or other pole of each conceptual scale. This result is patently ridiculous, and is due to considering the conceptual scales as if they were entirely separate. In fact, association of a sound with one concept implies its association to some degree with almost all other concepts in the language. This interlocking nature of the language ensures that extreme polarization seldom occurs, since polarization on some dimensions tends to depolarize other dimensions.

The overall resulting pattern of phonetic symbolism is the result of two opposed tendencies. Some sounds are strongly polarized on some conceptual scales and unpolarized on other scales. Some scales tend to have many polarized sounds; others, few. One may hazard a guess that cultural patterns may affect which scales tend to strong polarization. Taylor and Taylor (1962) suggested that cultural patterns play a part in determining which conceptual scales tend to show related patterns of polarization.

As a language evolves, and dialects diverge, so should the pattern of phonetic symbolism change. Where sound shifts occur, such as the shift from initial /t/ to /ts/ (Z) in High German, the phonetic symbolic values should follow the shift. Where an English observer finds initial /t/ to have some particular connotation, the German should find initial /ts/ to have the same connotation. The possibility arises that correlations of phonetic symbolism patterns in monolinguals might be used as a measure of historical relationship among languages. The idea is like that of determining the similarity of a set of words for basic concepts (Glottochronology; Hockett, 1958), except that no such arbitrary listing is required. The probability that cultural patterns affect the patterns of phonetic symbolism means that this type of measure must be severely tested before it is assumed to be reliable.

The analysis above has been largely in terms of initial sounds. The theory makes no distinction. Wherever associations may consistently recur between sounds and concepts, there phonetic symbolism will be found. The most prominent sounds in words are probably the initial sounds, the first vowel, and the stressed vowel. These points might be expected to contribute most to phonetic symbolism patterns, but only because of their prominence. Taylor and Taylor (1962) found no consistent difference in the strength of symbolism between initial and final consonants in CVC nonsense syllables. Of the 16 possible comparisons in 4 languages, 7 showed symbolism significant with $p < .001$ for both initial and final consonant, 3 showed the final consonant to give more significant symbolism than the initial consonant, and 6 showed the initial consonant more significant than the final.

While the theory does suggest why certain scales of meaning show more phonetic symbolism than others, and why the strongly symbolic scales vary from language to language, it does not give a reason why speakers of one language are more susceptible to phonetic symbolism than are speakers of another language. Both Miron (1960, 1961) and Taylor and Taylor (1962) found Japanese speakers giving far more strongly symbolic scores overall than did their English-speaking subjects. Taylor and Taylor also found Tamil speakers to give less variation in scores than English speakers give. These results occurred in spite of the fact that Japanese and Tamil speakers showed equal intersubject consistency, while English speakers showed more intersubject consistency than either. There seems to be a real effect. Different communities are differently susceptible to phonetic symbolism.

Conclusion

Though there have been numerous studies of the UPS hypothesis, there is still little direct evidence on several of the basic questions. However, the logical hierarchy among the different questions of the existence of phonetic symbolism suggests that the answers given in the evidence may be accepted with few qualms. Phonetic symbolism occurs in the words of a language, and it is detected in arbitrary phoneme clusters by speakers of a language. The symbolic connotations of a particular phoneme are the same in both cases. The symbolic connotations are not the same in the words of or between speakers of unrelated languages.

While the question of existence appears settled, the question of degree remains to be answered. For example, the results of Miron (1960, 1961) and Taylor and Taylor (1962) suggest that Japanese speakers are more susceptible to phonetic symbolism than are English speakers, and that within each group different dimensions show different strengths of symbolism. There is a suggestion (Newman, 1933, Taylor and Taylor, 1962) that the consistency of subjective phonetic symbolism may increase slightly with age. Such a trend would be anticipated by the language-dependent theory. Parametric studies of these and allied questions would go far toward clarifying the remaining areas of confusion.

REFERENCES

Allport, G. W. Phonetic symbolism in Hungarian words. Unpublished manuscript, Harvard University, 1935. (Cited by R. Brown, *Words and things.* Glencoe: Free Press, 1958.)

Brackbill, Yvonne, & Little, K. B. Factors determining the guessing of meaning of foreign words. *Journal of Abnormal and Social Psychology,* 1957, **54**, 313–318.

Broadbent, D. E., & Ladefoged, P. Vowel judgments and adaptation level. *Proceedings of the Royal Society of London, Series B,* 1960, **151**, 384–399.

Brown, R., Black, A. H., & Horowitz, Z. A. Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology,* 1955, **50**, 388-393.

Brown, R. *Words and things.* Glencoe: Free Press, 1958.

Dagiri, O. On phonetic symbolism. In Y. Endo, O. K. Hatano, H. Kobayashi, M. Koshimizu, Miyago, H. Nakajima, & T. Obonai (Eds.), *Science of language.* Vol. 3. Tokyo: Nakajima Shoten, 1958. (In Japanese)

Davis, R. The fitness of names to drawings: Cross-cultural study in Tanganyika. *British Journal of Psychology,* 1961, **52**, 259–268.

Eberhardt, Margarete. A study of phonetic symbolism of deaf children. *Psychological Monographs,* 1940, **52**(1, Whole No. 232).

Ertel, S., & Dorst, R. Expressive Lautsymbolik. *Zeitschrift für experimentelle und angewandte Psychologie,* in preparation.

Ertel, S. Der Lautcharakter künstlicher Lautgebilde. *Psychologische Forschung,* 1965, **28**, 491–578.

HOCKETT, C. F. *A course in modern linguistics.* New York: Macmillan, 1958.

HOUSEHOLDER, F. W. On the problem of sound and meaning, and English phonesthemes. *Word,* 1946, **2**, 83–84.

JESPERSEN, O. *Language.* New York: Holt, 1922.

LADEFOGED, P., & BROADBENT, D. E. Information conveyed by vowels. *Journal of Acoustical Society of America,* 1957, **29**, 98–104.

MALTZMAN, I., MORRISETT, L., & BROOKS, L. O. An investigation of phonetic symbolism. *Journal of Abnormal and Social Psychology,* 1956, **53**, 249–251.

MARCHAND, H. Phonetic symbolism in English word-formations. *Indogermanische Forschungen,* 1958–1959, **64**, 146–168, 256–277.

MCMURRAY, G. A. Meaning associated with the phonetic structure of unfamiliar foreign words. *Canadian Journal of Psychology,* 1960, **14**, 166–174.

MIRON, M. S. A cross-linguistic investigation of phonetic symbolism. Doctoral dissertation, University of Illinois. Ann Arbor, Mich.: University Microfilms, 1960. No. 60–3961.

MIRON, M. S. A cross-linguistic investigation of phonetic symbolism. *Journal of Abnormal and Social Psychology,* 1961, **62**, 623–630.

NEWMAN, S. S. Further experiments in phonetic symbolism. *American Journal of Psychology,* 1933, **45**, 53–75.

OSGOOD, C. E. Meaning. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics.* Bloomington: Indiana Univer. Press, 1965. P. 182.

ROSENTHAL, R. On the social psychology of the psychological experiment: The experimenter's hypothesis as an unintended determinant of experimental results. *American Scientist,* 1963, **51**, 268–283.

SAPIR, E. *Language.* New York: Harcourt, Brace, 1921.

SAPIR, E. A study in phonetic symbolism. *Journal of Experimental Psychology,* 1929, **12**, 225–239.

SAPORTA, S. Relations between psychological and linguistic units. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics.* Bloomington: Indiana Univer. Press, 1965. Pp. 60–65.

TAYLOR, INSUP K. Phonetic symbolism re-examined. *Psychological Bulletin,* 1963, **60**, 200–209.

TAYLOR, INSUP K., & TAYLOR, M. M. Phonetic symbolism in four unrelated languages. *Canadian Journal of Psychology,* 1962, **16**, 344–356.

TSURU, S., & FRIES, H. S. A problem in meaning. *Journal of General Psychology,* 1933, **8**, 281–284.

WEISS, J. H. Role of "meaningfulness" versus meaning dimensions in guessing the meanings of foreign words. *Journal of Abnormal and Social Psychology,* 1963, **66**, 541–546.

WEISS, J. H. Phonetic symbolism re-examined. *Psychological Bulletin,* 1964, **61**, 454–458.

ZIPF, G. K. *The psycho-biology of language.* Boston: Houghton Mifflin, 1935.

(Received February 25, 1965)

Psychological Bulletin
1965, Vol. 64, No. 6, 428-441

IS TIME-OUT FROM POSITIVE REINFORCEMENT AN AVERSIVE EVENT?

A REVIEW OF THE EXPERIMENTAL EVIDENCE[1]

HAROLD LEITENBERG[2]

Indiana University

Experiments concerning the aversive properties of time-out (TO) from positive reinforcement are reviewed. A discussion of experimental designs employed and problems encountered in defining aversive stimuli precedes the review. The major topics covered are: avoidance of TO, escape from TO, escape from stimuli which previously signalled TO, punishment with TO, effects on ongoing behavior of pre-TO stimuli, escape from conditions of positive reinforcement into TO. In general, TO satisfied sufficient criteria to conclude that it belongs to the class of stimuli called "aversive." However, since some differences between the effects of shock and TO were found; since some experiments allow other interpretations than that TO is aversive; and since so much research remains to be done, this conclusion is only tentative.

Withdrawing positive reinforcement and eliminating the opportunity to obtain positive reinforcement are commonly employed in our society as punishments for behavior judged to be undesirable. Sentence to prison, dismissal from work, loss of driver's license, and withheld allowance are a few examples. It is only in the last decade, however, that the aversive properties of the removal of positive reinforcement have begun to receive an experimental analysis of some depth.

The present paper reviews this experimental literature, with special emphasis put on those studies that represent different procedural approaches to the problem. A short discussion of what experimental operations might conceivably be subsumed by "time-out from positive reinforcement" precedes the review as does an equally brief summary of the usual defining criteria of an aversive event.

This paper's scope does not extend to a discussion of theories which suggest that omission of reward induces in the organism an emotional state labeled frustration (Brown & Farber, 1951, Lawson & Marx, 1958b). This hypothesized internal state and the cues associated with it are sometimes said to have specific aversive properties, as well as discriminitive, and diffuse energizing functions (e.g., Amsel, 1958, 1962; Marx, 1956).

Definition of TO from Positive Reinforcement

There is no single set of operations which adequately defines time-out (TO) from positive reinforcement. The essential feature is a period of time in which positive reinforcement is no longer available.

Within this boundary can fit such diverse experimental operations as extinction, S^Δ in discrimination training, and intervals between reinforcement in various schedules of reinforcement. These three procedures are distinguished in part by differences in the amount of stimulus overlap between periods of reinforcement and nonreinforcement. In extinction there are no cues which signal the absence of positive reinforcement; in S^Δ the absence of reinforcement is set off by a change in stimulus conditions, and in schedules of reinforcement the cues for periods of nonreinforcement might be time based or response produced. These procedures usually differ, also, in the duration and frequency of nonreinforced periods.

Studies which have used pigeons as subjects conventionally establish a blackout dur-

---

[1] This manuscript was prepared while the author had a predoctoral fellowship (1-Fl-MH-23, 283-01) from the National Institutes of Health. The author would like to thank James A. Dinsmoor for his helpful criticisms of an earlier draft of this paper.

[2] Now at the University of Vermont.

ing TO; all house lights and key lights are turned off. Since pigeons tend to be immobile in darkness, in effect the opportunity to respond is eliminated as well as the positive reinforcement. This could, therefore, be considered an example of blocking or interrupting an instrumental response chain which formerly led to food reinforcement.

Another possible way to arrange a TO from positive reinforcement is to block the consummatory response; reward can be visible but inaccessible. A barrier can separate the organism from the food, or the mouth can be muzzled to prevent eating but not contact. Reward can also be presented and withdrawn so quickly that eating is impossible.

There has not been sufficient work done in this area to show what different effects might be produced as a function of these variations in TO procedures. Different effects might be expected, however, simply because some procedures contain more elements of secondary reinforcement than others (e.g., Lawson & Marx, 1958a). Direct procedural comparisons need to be made within each of the paradigms used to define an aversive stimulus.

Definition of an Aversive Event

Discussion of what is meant by an "aversive" event will be restricted to the behavioral sphere because the only other current possibilities deal with subjective judgments of pain or unmeasured physiological disequilibrium. These are obviously fruitless. Aversiveness is a "hypothetical construct" and has no meaning independent of the defining experimental operations and observed behavioral effects.

An event is empirically defined as aversive if behavior which *prevents* or *terminates* it is maintained above operant level (Keller & Schoenfeld, 1950). This represents the avoidance and escape paradigms respectively. Another empirical definition of an aversive stimulus is that a response which *produces* it will decrease in frequency below its operant level. This is what Mowrer (1960) calls "passive avoidance" and Kimble (1961) calls "punishment training." A fourth possible defining property of an aversive stimulus is exemplified by the CER (conditioned emotional response) or conditioned suppression phenomenon (Estes & Skinner, 1941). Here, ongoing behavior maintained by a schedule of positive reinforcement is suppressed in the presence of a signal which systematically precedes and terminates coincident with the administration of electric shock. Finally, if a stimulus is a primary negative reinforcer, then neutral stimuli which set the occasion for its occurrence should themselves acquire secondary negative reinforcing properties.

One obvious problem with these different ways of measuring aversiveness is that they may not agree with each other. The explanation might simply be that they are not equally sensitive. Or, they may not be equivalent measures. It is an empirical question whether a stimulus meeting one criterion would meet any other. There doesn't seem to be any explicit basis, either, for considering one criterion "better" than another though a bias exists in the literature in favor of avoidance and escape learning.

Another difficulty is introduced by the fact that an event such as sudden stimulus change will satisfy for a short period of time the criterion measures of punishment training and conditioned suppression. Does this make stimulus change an aversive event? Or must duration of effect also be considered in the definition of a negative reinforcer? To further complicate matters, under certain test and training conditions stimulus change will act as a positive reinforcer (e.g., McCall, 1965).

Another ambiguity in the definition of aversive stimuli is presented by the case of satiation. It seems to satisfy the punishment criterion of an aversive event; if excessive amounts of reward are received, responding will be depressed. Note that repeated exposure to reward is response contingent, just as shock is response contingent in punishment training. Yet no one considers satiation to be aversive—unless its consequence is indigestion. This suggests that a complete definition of punishment requires further specification which would exclude the case of satiation.

It seems most reasonable, then, when trying to determine the aversive properties of a given event or stimulus, to use as many of these different defining techniques as possible. The more the measures agree with each other, the more certain it is that the event is aversive. This view is predicated on the relatively close agreement between measures when electric shock is the aversive stimulus in question, and it governs the manner in which the rest of this paper is organized.

Many of the studies reviewed in this paper were carried out in the operant tradition of repeated measurements on a small sample of subjects. The reliability of the data for the individual *S*s is, therefore, of a superior order than usual group data, but the generality of the reported finding is sometimes left in doubt.

Avoidance of TO from Positive Reinforcement

Experiment IV in Ferster's stimulating 1958 monograph illustrates one procedure for testing whether or not *S*s will learn to avoid TOs. Chimpanzees were employed as *S*s. Two levers located 12 in. apart were present simultaneously in the experimental enclosure. Responses on the right-hand lever were positively reinforced on a variable interval 3 min. schedule. A continuous avoidance schedule was programed on the left-hand lever. If no avoidance response occurred, a TO of 3 min. was scheduled every 45 sec. Each response on the left-hand lever, however, postponed the TO from the VI schedule for a certain duration. These durations, which Sidman calls response-stimulus (R-S) intervals in continuous shock avoidance (Sidman, 1953), were systematically varied: they were 1, 2, 5, and 10 min. Ferster found an increase in rate of avoidance responding as the interval was decreased.

Morse and Herrnstein (1956) reported a similar study, using pigeons as *S*s. Again two response alternatives were present concurrently. One response was maintained by a variable interval schedule of positive reinforcement, and the other response postponed TOs from this schedule. As with the chimps, response rate on the avoidance key increased as the R-S interval was decreased.

Thomas (1964) showed that when the mean interval between reinforcements was changed from 9 min. to 1 min., the number of TOs that were avoided decreased. Apparently when the frequency of reinforcement is higher, pigeons are more "willing" to tolerate TOs from the schedule of reinforcement. This is unexpected, because at short mean VI intervals, *S*s lose more reinforcements per TO than at long mean VI intervals. Avoidance of TO is possibly governed in part by some complex relationship between relative loss, absolute loss, and absolute frequency of reinforcement. This remains to be worked out.

Mechner and Ray (1959) demonstrated that rats will learn to avoid TOs from a fixed-interval 15-sec. schedule of reinforcement. In order to avoid a 2-min. interruption of this schedule, *S*s had to refrain from pausing longer than 3 sec. between successive responses on the food reinforced bar. When this contingency was removed, interresponse times became longer.

Baer (1960) investigated whether kindergarten-age children will learn to avoid having film cartoons cut off. He found an "escalator avoidance schedule" to be quite effective. In this procedure each avoidance response postponed the next film interruption in a cumulative manner. For instance, if the R-S interval was 5 sec. and *S* made 20 responses in rapid succession and then stopped responding, the cartoon would not be cut off until 100 sec. had elapsed from the first response. When Baer used the usual Sidman avoidance procedure—where each response resets anew the timer controlling the delay between response and TO—he failed to maintain avoidance behavior with this age group. This failure might have been caused by R-S intervals that were too short—3, 5, and 10 sec.—or because the Baer experiment, unlike the others, did not study avoidance of TOs from a schedule of response contingent positive reinforcers. Initially, *S*s were not required to make a response to produce the film cartoons.

The prevalent explanation of shock avoidance by reinforcement theorists (Dinsmoor, 1954; Schoenfeld, 1950; Sidman, 1953) is in terms of the immediate consequence of the avoidance response, namely, reduction of conditioned aversive stimuli. In Sidman avoidance, all behavior other than the avoidance response is sooner or later closely followed by shock, and therefore, tactile, proprioceptive, and environmental stimuli produced by these nonavoidance responses acquire aversive properties. The avoidance response is supposedly reinforced because it terminates these stimuli. This formulation could explain the data which illustrates avoidance of TO from positive reinforcement. It is only necessary to assume that response produced stimuli which precede the loss of positive reinforcement acquire secondary negative reinforcing properties. A later section this this paper reviews studies related to this assumption.

But another interpretation is possible—one that suggests that none of these TO avoidance studies provide clear evidence that TO from positive reinforcement is an aversive event. It could be said that these studies only indicate that *S*s will maintain behavior which leads to positive reinforcement. The concept of aversiveness seems unnecessary. Rather it seems adequate and more parsimonious to explain these findings just by saying: The pattern of behavior followed by most positive reinforcement is most strengthened. For in all the studies so far reviewed, *S* gains more positive reinforcement during a session if he makes the avoidance response than if he fails to make it.

Note, though, that in the studies cited, the avoidance response is often made on a different manipulandum from the one which leads to food reinforcement. Also some form of "change over delay" is usually arranged so that the avoidance response can never *immediately* precede the delivery of food and thereby receive adventitious positive reinforcement. For example, a response on the food key could be programed never to produce a reward if it followed within 2 sec. of an avoidance response.

Nonetheless, over a given period of time, avoidance behavior in conjunction with behavior on the food key will lead to the greatest amount of positive reinforcement. No other behavior pattern will do the same. And it is commonplace to note that behavior which successfully leads to the most positive reinforcement is maintained while other behavior is eliminated. If *S*s are given a choice between the right arm of a T maze in which reward is present on every trial, and the left arm of a T maze in which reward is present on only half of the trials, it is likely they would show a preference for the right arm. This does not make the left arm aversive; it just makes it less preferred. Few theorists would label choice of the right arm an avoidance response. Similarly, responding on both the avoidance bar and the food bar is the preferred response pattern because it leads to more reinforcement in a shorter time period.

This interpretation in no way negates the possibility that avoidance responding in these studies is maintained primarily because it postpones an aversive event—TO from positive reforcement. It is just unfortunate that the studies are not designed in a way which would make it possible to untangle the two interpretations.

What would be more convincing is a demonstration that *S*s will learn to avoid stimuli which previously set the occasion for TOs from positive reinforcement. Unlike studies reporting avoidance of TO itself, the avoidance response here would no longer have the added effect of increasing the amount of positive reinforcement *S* gets during a session, compared to what he would get if the avoidance habit had not been established.

Punishment Training with TO from Positive Reinforcement

In the punishment procedure, the aversive stimulus is response contingent, and attention is focused on the subsequent decline in the frequency of the punished response. The responses that take the place of the punished response—avoidance responses—are not usually examined.

Ferster has shown in a series of studies (Ferster, 1958, Exp. III; Ferster, 1960; Ferster, Appel, & Hiss, 1962) that chimps and pigeons will stop responding in the presence of a signal if responding leads to a TO, and if cessation of responding serves to avoid TO. A signal, such as a change in the color of the lighted disc at which the pigeon is pecking for food reinforcement, precedes a scheduled TO by a fixed duration. If the *S* stops pecking at the disc in the last 5 sec.

of the signal's presence, then the signal will terminate without being followed by a TO. If *S* continues to respond, a TO follows after the signal terminates. With pigeons, responding is almost completely suppressed during the entire time the signal is present. Chimps make a finer discrimination, continuing to respond until the critical period of the signal is reached.

This procedure should not be confused, as it sometimes seems to be, with the more familiar Estes-Skinner CER paradigm (1941). In the CER procedure, response suppression does not serve to avoid shock; changes in behavior in the presence of the CS have no programed consequence upon the delivery of shock. A later section of this paper reviews those studies which compare shock and TO in the CER paradigm.

Baer (1962) was able to suppress thumbsucking in nursery-school age children by interrupting film cartoons whenever the child had his thumb in the mouth. When the punishment contingency was withdrawn, thumbsucking reverted to its earlier frequency.

These punishment studies are as difficult to interpret as the avoidance studies were. The response contingent TOs reduce frequency of reinforcement per unit time. The higher positive reinforcement frequency maintained by the pattern of behavior that evolves (response suppression) seems sufficient explanation of the results. There is no need to postulate that the TOs were aversive in these studies.

It is possible to schedule contingent TOs (punishment with TO from positive reinforcement) in such a way that response suppression (should it occur) would not affect overall frequency of reinforcement.

Using mental hospital patients as *S*s, Holz, Azrin, and Ayllon (1963) demonstrated that TO from positive reinforcement has relatively weak suppressive powers when it is made contingent upon an operant response maintained at the same time by some positive reinforcement, in this case cigarettes. Every 10th response produced a TO of 30 sec. Cigarettes were delivered for the same response on a variable-interval schedule. In this procedure, response suppression, as long as it was not complete, neither increased nor decreased frequency of reinforcement. Two out of the four Ss showed a moderate decline in response frequency when the TO contingency was introduced. The other two *S*s showed no change in performance until a new response alternative was offered, one which continued to produce cigarettes on the same VI schedule but did not produce any TOs. Under these conditions all four *S*s stopped responding completely on the manipulandum leading to TOs and switched over to the new unpunished response.

The results of this study suggest that TO is, at best, a very poor punisher: shock would suppress responding before the second response alternative was introduced. However, it is worth noting that in the same paradigm loud noise had an effect more similar to TO than to shock (Herman & Azrin, 1964). The results of the Holz et al. study also support the hypothesis that TO acts like an aversive stimulus only when the behavior which satisfies the aversive criterion maintains a higher frequency of reinforcement than would be the case if it did not occur.

Matching to sample has been a popular technique for studying how TO affects behavor which produces it. Much of the work by Ferster and his associates has been done with pigeons. Birds must peck at a center key to produce a sample stimulus, usually a particular key color. After some fixed period, the light behind the center key is turned off and the two side keys are lit up. If the bird pecks the side key whose color corresponds to the former color of the center key, he is reinforced on some interval schedule. If he pecks the incorrect key, a TO is produced. A noncorrection procedure is employed so that once *S* makes his choice response on one of the side keys, both side key lights go off. If an incorrect response is made, the center key is relighted after the TO is over. Every correct response is reinforced with a brief flash of the magazine lights, while food is delivered on some schedule.

The major variables which have been investigated in the matching-to-sample studies are TO duration, the schedule by which TO is programed, and the schedule of reinforcement for correct matching responses. The first finding was that when every correct response was positively reinforced rather than intermittently reinforced, TOs following incorrect responses had little effect on matching accuracy (Ferster & Appel, 1961). As Ferster and Appel pointed out, interval schedules are generally more sensitive to independent variables than continuous reinforcement or ratio schedules. These authors found that when correct matching responses were reinforced on a VI 3-min. schedule, matching accuracy progressively improved as the TO duration for incorrect responses was raised from 1 to 10 to 30 to 60 sec. A TO duration of 2 min. disrupted matching performance by depressing both correct and incorrect responses indiscriminately.

Zimmerman and Ferster (1963) reported a similar finding with different TO durations. A 10-min. TO following incorrect responses impaired matching accuracy, while a middle range of TO durations improved matching accuracy. This study also showed that matching accuracy declined as the number of incorrect responses required to produce a TO was increased.

These matching-to-sample studies provide more substantial evidence than the other punishment studies that TO acts like an aversive stimulus. It is the generalized suppression of both incorrect and correct responses by long TOs that is important. This suppression does not have the added effect of increasing reinforcement frequency per unit time. In fact, response supression here leads to fewer reinforcements than if it did not occur.

Holder, Marx, Holder, and Collier (1957) published an interesting experiment which could be interpreted as illustrating the punishing effects of interrupting an instrumental response chain ending in positive reinforcement. Before they were allowed to complete their run to a goal box containing food, rats were delayed for different periods of time at a halfway point in a straight alley. The group that was delayed the longest took the longest time to run to the detention box. Possibly the running response was disrupted because it lead to the section of the runway associated with delay, a section which acquired aversive properties and provoked an approach-avoidance conflict.

Though the required length of a complete treatment is prohibitive for this paper, it seems desirable to at least briefly discuss the relevance of experimental extinction. The experimental operations of extinction of a positively reinforced response satisfy the punishment paradigm; the consequence of each formerly reinforced response is the omission of reward, and this paper assumes that omission of reward is a substantive stimulus. The question is whether or not it acts like other aversive stimuli in the punishment paradigm. In general, it does. The frequency of a formerly reinforced response declines when reward is withdrawn, just as it does when shock is introduced.

Although Amsel has not used the term punishment, his analysis of the inhibitory effects of nonreinforcement is in essence a punishment analysis. He suggests that absence of reward produces a motivational state of frustration which is chained back from the goal box and weakens the instrumental approach response. The approach response is weakened because it produces conditioned anticipatory frustration stimuli which are aversive. Consequently, responses which compete with the approach response are strengthened because they reduce this conditioned aversive frustration. This is the familiar avoidance explanation of punishment (Dinsmoor, 1954, 1955). In fact, Amsel says, "the greater the strength of frustration the faster the extinction (avoidance) of a nonrewarded response [1962, p. 308]." No doubt Dinsmoor would discard the concept of frustration as unnecessary, just as he discards the concept of fear when shock is the punishing stimulus. A more behaviorally oriented statement would be: because omission of reward is aversive the stimuli produced by the formerly reinforced response acquire aversive properties; therefore, other responses which terminate these conditioned aversive stimuli are strengthened at the expense of the formerly reinforced response.

Whether or not the avoidance hypothesis originally offered in explanation of the facts of punishment can account for the various phenomena of extinction is still unknown. It is certain, however, that the adequacy of such a theoretical account will depend a great deal on a finer experimental analysis of the similarities and differences between extinction (punishment with TO) and punishment (punishment with shock or other supposed aversive stimuli) than currently exists. For example, it would be interesting to know if spontaneous recovery and disinhibition occur with punishment as well as with extinction. To the author's knowledge, no such comparisons have ever been made.

Other comparisons have been made, however, and the evidence suggests that there are considerable similarities between the effects of extinction and punishment. A recent paradoxical finding that has been repeated quite often is that with a greater number of training trials there is less, not more, resistance to extinction (see Paul, 1965, for a related review). Similarly, it has been reported that as the number of training trials increases, the suppressive effects of punishment increase (Karsh, 1962; Miller, 1960).

Karsh (1964b), using a discrete trials operant procedure, studied the effects of intertrial intervals on the suppressive action of punishment. During the punishment phase of the experiment, *S*s undergoing 1 trial a day showed more resistance to the suppressive effects of punishment than *S*s receiving 10 or 50 trials a day. Although the results with extinction are not always consistent, the majority finding is that with spaced trials during extinction there is also greater resistance to extinction (Kimble, 1961, p. 283).

Reynolds observed "behavioral contrast" in pigeons with both response contingent shock and nonreinforcement. When responding in the presence of one stimulus was punished and suppressed with shock, rate of responding maintained by a VI schedule of reinforcement in the presence of another stimulus increased (Brethower & Reynolds, 1962). Similarly, when reinforcement for responding in the presence of one stimulus was withdrawn, rate of responding increased in the presence of another stimulus (Reynolds, 1961).

Brown and Wagner (1964), in a fascinating study, found that gradual exposure to shock during training increased resistance to extinction on later test trials and increased resistance to the suppressive effects of more intense shock. The converse was also true: occasional nonreinforced trials during training later increased both resistance to the suppressive effects of shock and resistance to the suppressive effects of extinction. This is a particularly clear demonstration of the similarity of shock and loss of reinforcement.

Likewise, Karsh (1964b) found that adding punishment to food reinforcement during a part of acquisition increased resistance to extinction. This finding is reminiscent of the partial reinforcement effect: interspersed nonreinforced trials during acquisition, as compared to continuous reinforcement during acquisition, increases resistance to extinction. It would be useful to make a parametric investigation of the effects of different percentages of punishment trials during acquisition on resistance to extinction.

Some drugs produce similar effects on responding suppressed by shock and responding suppressed by experimental extinction. Sodium amytal and alcohol lessen the degree of response suppression during approach-avoidance conflict periods (Barry & Miller, 1962). These two drugs also lessen the degree of response suppression caused by the withdrawal of reinforcement during extinction (Barry, Wagner, & Miller, 1962).

Fewer experiments demonstrating differences between extinction and punishment were found. Holz and Azrin (1963) suggested that extinction suppresses behavior less quickly and less completely than moderate levels of shock, and that when conditions of extinction and punishment are withdrawn and reinforcement is reintroduced, responding recovers more rapidly after extinction than after punishment. As Holz and Azrin suggested, these differences may simply be due to differences in *degree* of aversiveness between response-contingent extinction and response-contingent shock. Other differences are (*a*) A greater magnitude of reward during training produces less resistance to extinction (Amsel, 1962, cites several studies); while the converse relationship is found between size of reward and resistance to punishment (Bower & Miller, 1960). This is, perhaps, not surprising because there is probably a relationship between the size of the reward that is withdrawn and the degree of aversiveness of extinction, while no such relationship exists between the size of reward and degree of aversiveness of shock. Further tests of this hypothesis need to be made in aversive paradigms other than punishment. (*b*) The total number of responses emitted during extinction was unchanged even though an initial period of mild punishment temporarily suppressed responding (Estes, 1944), suggesting that the two procedures control response frequency via different processes.

Escape from TO and Escape from Stimuli Associated with TO

Any study which shows operant behavior being maintained by a schedule of positive reinforcement can be interpreted as demonstrating escape from periods of nonreinforcement. Responding in the intervals between reinforcements can be said to serve two functions at the same time—terminating a period of nonreinforcement and leading to the receipt of positive reinforcement. This is, obviously, not very satisfactory evidence that TO from positive reinforcement has negative reinforcing properties.

A study in which the escape response did not simultaneously lead to positive reinforcement and to termination of the supposed aversive stimulus was done by Adelman and Maatsch (1956). Three groups of 10 rats each were trained to jump 10 in. to a perch surrounding the top of a goal box. The goal box was at the end of a straight alley. The reward for jumping out of the goal box varied for the three groups. For one group the goal box had previously contained food reward for running up the alley. Two other groups without this experience received equal amounts of direct placements into an empty goal box before the jump-out phase of the experiment was introduced. The first group's supposed reward for jumping out of the goal box was escape from an aversive frustrating situation, that is, an empty goal box in place of one with food. The second group received a pellet every time they jumped out of the box and onto the surrounding perch. The only possible reinforcement for the third group was escape from confinement and/or satisfaction of some exploratory drive. The results clearly demonstrated that when the box was associated with the removal of food, *S*s jumped fastest and continued to perform at the same level for over 100 trials.

This is a somewhat surprising finding. Ordinarily, one would expect a stimulus associated many times with reward to become a secondary reinforcer and if it is never again paired with the primary reinforcer to lose its effectiveness as a secondary reinforcer *gradually*. Then it would revert back to a neutral stimulus or, perhaps, as this study suggests, become an aversive stimulus. It is not surprising that the stimulus becomes aversive, but it is surprising that the stimulus changes its function so rapidly. Although *S*s continue to run to the empty box, once they get there and find no reward they escape. Perhaps being kept in the empty goal box for 5 min. before being allowed to escape is a critical factor.

More recently, Wagner (1963) has shown that rats will learn to jump over a hurdle to terminate stimuli associated with the absence of reward. Two groups of rats equated for weight were trained to run down an alley on a 50% partial reinforcement schedule. For the experimental group, an auditory or visual CS was presented in the alley and in the goal box on nonreinforced trials. For the control group, the CS was presented in the retaining cage and did not signal nonreinforcement. In test trials, experimental *S*s jumped over a hurdle faster than control *S*s when this response served to terminate the CS. As training progressed, escape latencies decreased for the experimental group while they increased sightly for the control group.

These two studies are crucial for two reasons. First, the best evidence so far that TO from positive reinforcement is aversive comes from studies done in the punishment paradigm, but convention has it that escape and avoidance are more important criteria of aversive stimuli. As suggested earlier, previous experiments trying to demonstrate avoidance and escape from loss of reinforcement had a flaw. It was not clear whether the avoidance behavior—which was successfully demonstrated—was maintained because an aversive event was thereby postponed or because a greater frequency of reinforcement was thereby assured. Escape behavior in the Adelman and Maatsch experiment and in the Wagner experiment, on the other hand, does not appear to have any other reinforcement than the removal of what must be considered conditioned aversive stimuli. To insure the reliability of this phenomenon, however, it would be useful to have a few more examples.

Secondly, these two studies attest to the validity of the major assumption of the usual explanation of avoidance behavior. For both shock avoidance and TO avoidance, the assumption is that neutral stimuli paired with the stimulus to be avoided will acquire secondary negative reinforcing power.

This is the last evidence that the author will review which strongly supports the notion that TO is an aversive stimulus. The remaining experiments could be considered negative evidence; however, "qualifying evidence" seems the more appropriate term.

Effects of Pre-TO Stimuli

Ferster (1958, Exp. I) has published some data concerning the effects on ongoing behavior of a signal which precedes a TO from a schedule of reinforcement. He employed a variant of the Estes and Skinner (1941) CER paradigm which uses shock as the aversive stimulus.

In the typical CER procedure, stable reference behavior is maintained by a variable-interval schedule of reinforcement. Periodically during the session, a CS is presented for a fixed duration, and coincident with its termination a shock is delivered. Since the reinforcement schedule continues in effect while the CS is present, reward is potentially obtainable in the presence of the CS. Nonetheless, the repeated finding has been that ongoing behavior maintained by a schedule of positive reinforcement is suppressed in the presence of preshock stimuli (see Annau & Kamin, 1961; Brady & Hunt, 1955; Hoffman & Fleshler, 1961; Hunt & Brady, 1955; Lyon, 1963; Stein, Sidman, & Brady, 1958). The CS tends to elicit conditioned fear responses such as freezing and defecation which supposedly interfere with the ongoing reinforced behavior (e.g., Hunt & Brady, 1955).

In Ferster's study, the first response after the CS had been on for 30 sec. produced the TO. (Note that this procedure is decidedly different from the response-independent CER paradigm. In fact, it could be considered more representative of the punishment procedure than of anything else.) In any case, Ferster found that for chimps, response rates accelerated above control levels in the presence of the pre-TO stimulus.

Herrnstein (1955), in an unpublished thesis, employed the same response contingent procedure and reported similar findings with pigeons, but only when the VI schedule of reinforcement interrupted by the TO had relatively long mean intervals between reinforcements, such as 7 and 9 min. Why he failed to find the conditioned acceleration when the VI schedules had shorter mean intervals is still unknown. More work is needed to clarify the relationship of frequency of reward, nature of schedule, response rate, and the effects of the pre-TO CS. Herrnstein also found that as TO duration was increased, conditioned acceleration in the presence of the pre-TO stimulus was increased. The range of TO durations investigated was from 30 sec. to 50 min., and the schedule of reinforcement remained the same VI 7 min. for each duration.

An unpublished thesis by Leitenberg (1965) employed a multiple-schedule design to compare in the same pigeon the effects on ongoing behavior of preshock stimuli and pre-TO stimuli. Instead of the usual blackout, only the key light was turned off during TO. Different colored key lights signalled shock and TO. The comparison was made with four birds, using both the usual CER response independent procedure and the response contingent procedure which both Herrnstein and Ferster used in their TO studies. It made no difference which procedure was in effect: rate of pecking at a key on a VI 2.25-min. schedule accelerated in the presence of the pre-TO stimulus and declined in the presence of the preshock stimulus. With very weak shock intensities, response rate accelerated slightly in the preshock stimulus; however, this acceleration was apparently only a function of generalization because it was still present when there was no shock, and it disappeared when the TOs were removed. Clearly the effects on ongoing behavior of stimuli which signal TO are different from the effects of stimuli which signal shock—even when extremely low shock intensities are employed. The difference is not simply in degree of effect but in the direction of effect; pre-TO stimuli accelerate ongoing behavior while preshock stimuli suppress ongoing behavior.

What is the basis for this difference? Both the CS that precedes shock and the CS that precedes TO acquire negative reinforcing properties. Perhaps the preshock CS acquires greater aversive properties than the pre-TO CS. With relatively intense shock, when the key light changed to the color signaling shock, pigeons were observed to turn their backs to the key. This kind of escape response could readily account for the suppression of key-pecking behavior in the presence of the preshock CS. (Unlike rats, pigeons were not observed to freeze or defecate upon onset of the preshock CS, so these emotional responses could hardly account for the disruption in ongoing behavior.) The pigeons did not turn from the key in the presence of the pre-TO CS because it was not aversive enough to counteract the positively reinforced key-pecking behavior. A more formal test of this hypothesis is desirable. *Ss* will make a response to terminate a CS that precedes shock even if the shock is unavoidable (Kamin, 1956). Will they do the same with TO?

But this analysis does not explain why key-pecking frequency accelerated in the presence of the pre-TO CS; it just explains why it did not decelerate. The unconditioned response to omission of reward is heightened activity or heightened emotionality. Amsel and his co-workers have shown in a host of studies (Amsel, 1958, 1962) that running speed following nonreinforcement is faster than running speed following reinforcement, and faster than running speed following detention in a box where reinforcement was never before received. In the early periods of extinction of a positively reinforced response, bursts of rapid responding have been noted (Thompson, 1961, 1962). It has also been found that rats press the bar with greater force during extinction than during acquisition (Notterman, 1959). Wagner (1963) found that a startle response to a loud 1,000-cps tone was more vigorous when the tone was immediately preceded by a CS previously associated with nonreinforced trials in a partial reinforcement experiment.

Although this is pure speculation, it is possible that the energizing effects produced by TO were conditioned to the pre-TO CS. This conditioned "excitement" acts upon the most strongly established habit, key-pecking behavior, to cause an increase in rate of response.

The fact that a preshock CS suppresses ongoing behavior while a pre-TO CS accelerates ongoing behavior does not mean that TO is not an aversive stimulus. All it might signify is that TO does not act like shock in this particular paradigm—a paradigm which perhaps should not be used to define an aversive stimulus.

Interestingly enough, when loud noise was used, neither conditioned suppression as with shock nor conditioned acceleration as with TO was observed (Leitenberg, 1965). Whether or not other aversive stimuli, such as bright light, extreme temperatures, painful blows, or pinches will lead to conditioned suppression, like shock; or conditioned acceleration, like TO; or have very little effect, like loud noise; is still unknown.

Escape into TO from Positive Reinforcement

Recently, a series of studies demonstrated that rats and pigeons will impose upon themselves interruptions from fixed-ratio schedules of positive reinforcement (Appel, 1963; Azrin, 1961; Thompson, 1964; Zimmerman & Ferster, 1964).

Essentially the same procedure has been used in all these studies. Two response alternatives are available to *S*. Responses on one key are reinforced on a fixed-ratio schedule of reinforcement. The first response on the other key produces a TO. The second response on this switching key reinstates the fixed-ratio schedule on the food key. In this way the duration as well as the initiation of the TO is under the organism's own control. Zimmerman and Ferster (1964) presented data which indicates that if the TO duration is fixed at 10 sec. rather than being under the *S*'s own control, fewer TO-producing responses will be made.

Azrin's (1961) and Appel's (1963) studies with pigeons show that as the number of responses required for reinforcement is raised, more TOs are produced, and they are kept in force for a longer time. The range of values in Azrin's study was from FR 60 to FR 200, and the range in Appel's study was FR 80 to FR 240. At FR 200 Azrin noted that almost 50% of the session was spent in TO conditions, and at FR 240 Appel found that 30% of the session was spent in TO conditions.

Thompson (1964) showed that rats will press a bar three times in succession in order to produce 30-sec. TOs from fixed-ratio schedules of reinforcement. More TOs were produced as the fixed-ratio requirement for water reinforcement was raised, and fewer TOs were produced when the ratio requirement was systematically lowered.

Appel (1963) has also shown that when responses on the switching key are no longer effective in producing TOs, responding on this key undergoes the typical decline of extinction.

A fixed-ratio schedule of reinforcement produces a characteristic pause in responding just after each food delivery. Following this postreinforcement pause, *S*s usually rapidly run off the required number of responses without any systematic pause occurring again until after the next reinforcement is received. Appel (1963), Azrin (1961), and Thompson (1964) observed that most TO producing responses occurred during these postreinforcement pauses between ratio runs.

Zimmerman and Ferster (1964) found that the more different the stimulus conditions are during the TO from what they are during the time the schedule is in effect, the more TOs the pigeon will impose upon himself. The direction of the stimulus change associated with the onset of the TO does not, however, seem important. Thompson (1964) with rats and Azrin (1961) with pigeons found no difference between TOs introducing a decrease in illumination in the box and TOs introducing an increase in illumination.

These studies do not preclude the possibility that the change in stimulus conditions per se is the major reinforcement for TO producing responses (cf. Appel, 1963). The prevalent view, however, is that the response leading to TO is an escape response; it terminates stimulus conditions associated with a schedule which is supposed to have aversive properties—either because of the amount of work required for positive reinforcement or because of the long time intervals which elapse between successive reinforcements. What leads one to believe that escape is involved here is the increasing tendency to produce a change in stimulus conditions as the fixed-ratio requirement is raised.

It would be interesting to know if *S*s would escape from other difficult schedules of reinforcement such as a DRL schedule which requires *S*s to pause for certain lengths of time between successive responses, and a fixed-interval schedule with very long intervals between successive reinforcements. A yoked comparison between a variable-interval schedule and a high fixed-ratio schedule would be especially useful. A yoked procedure could equate the number of reinforcements and the intervals between reinforcements, but a large number of responses would not be required or made prior to each reinforcement by *S*s run under the interval schedule. In this way the factors responsible for the escape behavior—number of responses, or interval between reinforcement, or both—can be determined.

Allowing *S*s to interrupt the schedule and impose rest periods upon themselves is an excellent way to study the aversive properties of fatigue in animals. The FR requirement could be kept constant, while the amount of force necessary to press the bar could be varied systematically. It is expected that more escape responses would be made as the force requirement was raised on the food-contingent bar.

Experiments demonstrating escape into TO from a schedule of positive reinforcement indicate that: (*a*) some schedules of reinforcement can contain both positive and negative reinforcement properties; (*b*) depending on the nature of the underlying reinforcement schedule, removal of the opportunity to acquire reinforcement must now be considered potentially either a positive or negative reinforcer for nonsatiated organisms.

Concluding Remarks and Summary

The title of this paper asks what appears to be a simple question: is TO from positive reinforcement an aversive event? Upon examination, however, the question becomes somewhat complex. There is no single TO procedure, and there is no single experimental procedure for defining an aversive stimulus. Thus, there can hardly be a simple yes or no answer.

The use of TO in five different paradigms was analyzed: avoidance, escape, punishment, conditioned or secondary negative reinforcement, and CER. Expectations about what effects TOs would have in these paradigms (if indeed, TO were aversive) were mainly determined by the effects that electric shocks have had in previous studies. Although there have been some experiments employing other supposed aversive stimuli, such as loud noise, bright light, extreme temperatures, physical blows and pinches, and strange objects, they are greatly outnumbered by the literature on electric shock. Therefore, the first general prescription for future work in this area is: compare TO with a *variety* of other aversive stimuli in as many different experimental paradigms as possible. Shock may not be as representative of the class of aversive stimuli as is generally assumed; and some paradigms may yield similarities between stimuli, while others yield differences.

Although it can be said that in general TO satisfies the major criteria of an aversive stimulus, the results of the reviewed experiments are by no means definitive. The experimental evidence presents a complex picture of the aversive properties of TO from positive reinforcement. However, complexity is not automatically cause for despair. It is reliability and validity that matter, not simplicity. The major findings of this review were: (*a*) The design of studies demonstrating avoidance and escape from TO allow another interpretation than that TO is aversive. Avoidance and escape behavior could have been maintained because they led to a greater positive reinforcement frequency over a given time period than did the absence of such behavior. (*b*) A similar problem of interpretation was inherent in some of the punishment studies. The other punishment studies, however, provide reasonably adequate evidence that TO is an effective punishing stimulus, that is, it suppresses behavior which produces it; and that experimental extinction of a formerly reinforced response could be considered a special case of punishment. Although the procedures and effects of punishment and extinction were shown to have much in common, much more experimental work and theoretical analysis needs to be done before a final conclusion can be reached. (*c*) The most convincing evidence that TO is aversive comes from those studies demonstrating escape from stimuli which previously set the occasion for nonreinforcement. (*d*) Evidence implying differences between TO and other aversive stimuli include those studies which show that ongoing positively reinforced behavior is suppressed in the presence of stimuli preceding shock, but is accelerated in the presence of stimuli preceding TO from positive reinforcement. This difference was assumed to be a function of two factors: since pre-TO visual stimuli acquire less aversive properties than preshock stimuli, *S*s are less likely to turn away from them; and stimuli preceding TO elicited "conditioned excitement" which acted upon the prevailing key-pecking habit to cause accelerated response rates. (*e*) Studies illustrating self-imposed interruptions from high fixed-ratio schedules of positive reinforcement indicate that, depending on the nature of the schedule, TO could be either a positive or negative reinforcer.

In addition to analyzing implications of existing results, this paper also advanced specific proposals for new research. In general, future research should be directed mainly towards finding independent variables having a functional relationship with the aversive properties of TO. It should not be directed solely towards giving yes or no answers to the question: does TO have aversive properties?

Continuing research in this area should have great influence on theories of extinction, discrimination learning, and delay of reinforcement. In addition, increased experimental evidence relating to the punishing and negative reinforcing properties of removing reward, delaying reward, preventing work for reward, blocking consummatory behavior, and even reducing magnitude of reward should be of value to all those interested in the aversive control and modification of behavior. If Solomon (1964) is correct in saying that the area of aversive control has been relatively neglected, it is also correct to say that these TO variables have been severely neglected in our research activity—a neglect which makes no sense in terms of theoretical relevance or in terms of practical use in our society.

REFERENCES

Adelman, H. M., & Maatsch, J. L. Learning and extinction based upon frustration, food reward, and exploratory tendency. *Journal of Experimental Psychology,* 1956, **52**, 311–315.

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin,* 1958, **55**, 102–119.

Amsel, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review,* 1962, **69**, 306–328.

Annau, Z., & Kamin, L. J. The conditioned emotional response as a function of intensity of the US. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 428–432.

Appel, J. B. Aversive aspects of a schedule of positive reinforcement. *Journal of the Experimental Analysis of Behavior,* 1963, **6**, 423–428.

Azrin, N. H. Time-out from positive reiforcement. *Science,* 1961, **133**, 382–383.

Baer, D. M. Escape and avoidance responses of preschool children to two schedules of reinforcement withdrawal. *Journal of the Experimental Analysis of Behavior,* 1960, **3**, 155–159.

Baer, D. M. Laboratory control of thumbsucking by withdrawal and representation of reinforcement. *Journal of the Experimental Analysis of Behavior,* 1962, **5**, 525–528.

Barry, H., III, & Miller, N. E. Effects of drugs on approach-avoidance conflict tested repeatedly by means of a "telescope alley." *Journal of Comparative and Physiological Psychology,* 1962, **55**, 201–210.

Barry, H., III, Wagner, A. R., & Miller, N. E. Effects of alcohol and amobarbital on performance inhibited by experimental extinction. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 464–468.

Bower, G. H., & Miller, N. E. Effect of amount of reward on strength of approach in an approach-avoidance conflict. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 59–62.

Brady, J. V., & Hunt, H. F. An experimental approach to the analysis of emotional behavior. *Journal of Psychology,* 1955, **40**, 313–324.

Brethower, D. M., & Reynolds, G. S. A facilitative effect of punishment on unpunished behavior. *Journal of the Experimental Analysis of Behavior,* 1962, **5**, 191–199.

Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables with suggestions toward a tentative theory of frustration. *Psychological Bulletin,* 1951, **58**, 465–495.

Brown, R. T., & Wagner, A. R. Resistance to punishment and extinction following training with shock or nonreinforcement. *Journal of Experimental Psychology,* 1964, **68**, 503–507.

Dinsmoor, J. A. Punishment: I. The avoidance hypothesis. *Psychological Review,* 1954, **61**, 34–46.

Dinsmoor, J. A. Punishment: II. An interpretation of empirical findings. *Psychological Review,* 1955, **62**, 96–104.

Estes, W. K. An experimental study of punishment. *Psychological Monographs,* 1944, **57**(3, Whole No. 263).

Estes, W. K., & Skinner, B. F. Some quantitative properties of anxiety. *Journal of Experimental Psychology,* 1941, **29**, 390–400.

Ferster, C. B. Control of behavior in chimpanzees and pigeons by time-out from positive reinforcement. *Psychological Monographs,* 1958, **72**(8, Whole No. 461).

Ferster, C. B. Suppression of a performance under differential reinforcement of low rates by a pre-time-out stimulus. *Journal of the Experimental Analysis of Behavior,* 1960, **3**, 143–153.

Ferster, C. B., & Appel, J. B. Punishment of S^{Δ} responding in matching to sample by time-out from positive reinforcement. *Journal of the Experimental Analysis of Behavior,* 1961, **4**, 45–56.

Ferster, C. B., Appel, J. B., & Hiss, R. A. The effect of drugs on a fixed ratio performance suppressed by a pre-time-out stimulus. *Journal of the Experimental Analysis of Behavior,* 1962, **5**, 73–88.

Herman, R. L., & Azrin, N. H. Punishment by noise in an alternative response situation. *Journal of the Experimental Analysis of Behavior,* 1964, **7**, 185–188.

Herrnstein, R. J. Behavior consequences of the removal of a discriminative stimulus associated with variable interval reinforcement. Unpublished doctoral dissertation, Harvard University, 1955.

Hoffman, H. S., & Fleshler, M. Stimulus factors in aversive control: The generalization of conditioned suppression. *Journal of the Experimental Analysis of Behavior,* 1961, **4**, 371–378.

Holder, W. B., Marx, M. H., Holder, E. E., & Collier, G. Response strength as a function of delay of reward in a runway. *Journal of Experimental Psychology,* 1957, **53**, 316–323.

Holz, W. C., & Azrin, N. H. A comparison of several procedures for eliminating behavior. *Journal of the Experimental Analysis of Behavior,* 1963, **6**, 399–412.

Holz, W. C., Azrin, N. H., & Ayllon, T. Elimination of behavior of mental patients by response-produced extinction. *Journal of the Experimental Analysis of Behavior,* 1963, **6**, 407–412.

Hunt, H. F., & Brady, J. V. Some effects of punishment and intercurrent "anxiety" on a simple operant. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 304–310.

Kamin, L. J. The effects of termination of the CS and avoidance of the US on avoidance learning. *Journal of Comparative and Physiological Psychology,* 1956, **47**, 65–72.

Karsh, E. B. Effects of number of rewarded trials and intensity of punishment on running speed. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 44–51.

Karsh, E. B. Punishment: Effect on learning and resistance to extinction of discrete operant behavior. *Psychonomic Science,* 1964, **1**, 139–140. (a)

Karsh, E. B. Punishment: Trial spacing and shock intensity as determinants of behavior in a discrete operant situation. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 299–302. (b)

Keller, F. S., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

Lawson, R., & Marx, M. H. A comparison of some presumed frustrating and secondary reinforcing operations. *Journal of Comparative and Physiological Psychology,* 1958, **51**, 742–746. (a)

Lawson, R., & Marx, M. H. Frustration: Theory and experiment. *Genetic Psychology Monographs,* 1958, **57**, 393–464. (b)

Leitenberg, H. Conditioned acceleration versus conditioned suppression in the pigeon. Unpublished doctoral dissertation, Indiana University, 1965.

Lyon, D. O. Frequency of reinforcement as a parameter of conditioned suppression. *Journal of the Experimental Analysis of Behavior,* 1963, **6**, 95–98.

Marx, M. H. Some relations between frustration and drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1956.* Lincoln: Univer. Nebraska Press, 1956. Pp. 92–130.

McCall, R. B. Stimulus change in light-contingent bar pressing. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 258–262.

Mechner, F., & Ray, R. Avoidance of time out from fixed-interval reinforcement. *Journal of the Experimental Analysis of Behavior,* 1959, **2**, 261.

Miller, N. E. Learning resistance to pain and fear: Effects of overlearning, exposure, and rewarded exposure in context. *Journal of Experimental Psychology,* 1960, **60**, 137–145.

Morse, W. H., & Herrnstein, R. J. The maintenance of avoidance behavior using the removal of a conditioned positive reinforcer as the aversive stimulus. *American Psychologist,* 1956, **11**, 430. (Abstract)

Mowrer, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

Notterman, J. M. Force emission during bar pressing. *Journal of Experimental Psychology,* 1959, **58**, 341–347.

Paul, C. Effects of overlearning upon single habit reversal in rats. *Psychological Bulletin,* 1965, **63**, 65–72.

Reynolds, G. S. Behavioral contrast. *Journal of the Experimental Analysis of Behavior,* 1961, **4**, 57–71.

Schoenfeld, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. H. Hoch & J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950. Pp. 70–99.

Sidman, M. Two temporal parameters of the maintenance of avoidance behavior by the white rat. *Journal of Comparative and Physiological Psychology,* 1953, **46**, 253–261.

Solomon, R. L. Punishment. *American Psychologist,* 1964, **19**, 239–253.

Stein, L., Sidman, M., & Brady, J. V. Some effects of two temporal variables on conditioned suppression. *Journal of the Experimental Analysis of Behavior,* 1958, **1**, 153–162.

Thomas, J. R. Avoidance of time-out from two VI schedules of positive reinforcement. *Journal of the Experimental Analysis of Behavior,* 1964, **7**, 168.

Thompson, D. M. Escape from S^{Δ} associated with fixed-ratio reinforcement. *Journal of the Experimental Analysis of Behavior,* 1964, **7**, 1–8.

Thompson, T. Effect of chloropromazine on "aggressive" responding in the rat. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 398–400.

Thompson, T. The effect of two phenothiazines and a barbiturate on extinction-induced rate increase of a free operant. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 714–718.

Wagner, A. R. Conditioned frustration as a learned drive. *Journal of Experimental Psychology,* 1963, **66**, 142–148.

Zimmerman, J., & Ferster, C. B. Intermittent punishment of S^{Δ} responding in matching to sample. *Journal of the Experimental Analysis of Behavior,* 1963, **6**, 349–356.

Zimmerman, J., & Ferster, C. B. Some notes on time-out from positive reinforcement. *Journal of the Experimental Analysis of Behavior,* 1964, **7**, 13–19.

(Received April 29, 1965)

Psychological Bulletin
1965, Vol. 64, No. 6, 442–452

THREE-MODE FACTOR ANALYSIS[1]

JOSEPH LEVIN

University of Illinois[2]

2 studies employing Tucker's 3-mode factor analysis are reported. The 1st is an analysis of semantic differential data. 4 scale factors and 4 concept factors were obtained. 1 subject-type was obtained. It consisted of a core matrix linking the scale factors to the concept factors. The 2nd study is an analysis of an S-R Inventory of Anxiousness. 3 situation factors and 3 response factors were obtained. Also, 3 types were obtained, each consisting of a 3 × 3 matrix linking the response factors to the situation factors.

Classical factor analytic procedures are applicable to two-way classification data; for example, to matrices of subjects by tests. Assuming that the data have already been standardized properly, the data matrix can be factored into a product of two matrices: a factor score matrix and a factor loading matrix. However, psychological data are often three-way (or multi-way) classification data; for example, multi-trait multi-method matrices (Campbell & Fiske, 1959), the semantic differential (Osgood, Suci, & Tannenbaum, 1957), and the S-R Inventory of Anxiousness (Endler, Hunt, & Rosenstein, 1962).

The problem of analyzing multi-mode data (i.e., multi-way classification tables) has been discussed in the literature frequently (Abelson, 1960; Burt, 1955; Cattel, 1952; Guttman, 1958; Mahmoud, 1955). Yet all the proposed solutions are based on some sort of reduction of the data to a two-way table. The semantic differential is frequently reduced to a concept by scale table, averaging over subjects. Whenever individual differences are considered important, and the mode of "subjects" cannot be eliminated, the items are summed (or averaged) over the "unimportant" mode; for example, data given in a table of subjects by tests by occasions can be reduced by averaging over occasions. The S-R Inventory of Anxiousness (Endler, Hunt, & Rosenstein, 1962), which consists of responses to situations by subjects, can be reduced to a response by subject table, averaging each response over all situations. The latter is an example where the asymmetrical treatment, involved in most reduction procedures, is not tenable. Indeed, in their analysis of the S-R Inventory, Endler et al. have perceived this point and have treated the data symmetrically. Summing situations, they obtained a subject by response matrix. Summing responses, they obtained a subject by situation matrix. They also pointed out the problem of the relation among the factors obtained from each of these matrices, respectively.

The problem of a simultaneous analysis of all the modes of the data and of the relation among the corresponding factors is the main issue of three-mode factor analysis.

The Logic of Three-Mode Factor Analysis

The mathematical problem of three-mode factor analysis has been solved by L. R Tucker[3] (Tucker, 1964). The following section is an exposition of the structure and logical meaning of this method.

Three-mode factor analysis is not a straightforward generalization of classical two-mode factor analysis. Two-mode factor analysis requires some modification before it can be generalized. It is, therefore, simpler, first to introduce a modified version of two-mode factor analysis and explain three-mode factor analysis as a straightforward generalization. Besides, this modfication also has intrinsic interest in the two-mode case.

---

[1] This paper is based on a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Illinois.

The author expresses his acknowledgment to Ledyard R Tucker for his invaluable help in direction of the thesis. He also thanks Edward Ware and J. McV. Hunt for the data they made available for reanalysis.

This research received partial support from a contract with the Office of Naval Research, Contract No. US NONR 1834(39), Ledyard R Tucker, Principal Investigator. This support is gratefully acknowledged.

[2] Now at the Szold Foundation and the University of Tel-Aviv, Israel.

[3] Implication of factor analysis of three-way matrices for measurement of change. Paper presented at the Conference in Measuring Change, University of Wisconsin, May 1962.

A Modification of Two-Mode Factor Analysis

Let us reduce the semantic differential data (subjects by scales by concepts) to a two-way classification averaging over subjects. We obtain a scales by concepts matrix, X, of order $i \times k$ and of rank $p (p \leq i,\ p \leq k)$. Let us assume that the scores have already been standardized with regard to origin and unit of measurement, so that sum of products can be used instead of correlations. With these assumptions, the Eckart-Young procedure can be employed (Eckart & Young, 1936). According to this procedure ${}_iX_k$ can be factored as a product of three matrices:

$$ {}_iX_k = {}_iU_i\Lambda_k V_k, \qquad [1]$$

where $U =$ the orthonormal latent vector matrix of XX^T,
$V =$ the orthonormal latent vector matrix of X^TX,
$\Lambda =$ a rectangular matrix with 0 entries off diagonally, p nonzero entries $(\lambda_1 \cdots \lambda_p)$ in the diagonal starting with the upper left corner, and λ_j^2 $(j = 1 \cdots p)$ are the latent roots of XX^T (or of X^TX). Without loss of generality it may be assumed that the λ_j are arranged in decreasing order of magnitude.

If the m largest values of Λ are chosen, then

$$ {}_i\ddot{X}_k = {}_iU_m\Lambda_m V_k \qquad [2]$$

is a least square approximation to X for m factors. This can be reformulated so that U and V become principal-axes factor loading matrices:

$$\begin{aligned} \ddot{X} &= (U\Lambda)\Lambda^{-1}(\Lambda V) \\ &= \bar{U}\Lambda^{-1}\bar{V}; \end{aligned} \qquad [3]$$

$\bar{U}$ and $\bar{V}$ are the factor loading matrices for scales and concepts, respectively. The next step is to introduce a rotation to simple structure. Due to the logical symmetry of concepts and scales we have to rotate both $\bar{U}$ and $\bar{V}$.

Let $\bar{U}$ be rotated by T, and $\bar{V}$ be rotated by P

$$\ddot{X} = (\bar{U}T)T^{-1}\Lambda^{-1}P^{-1}(PV). \qquad [4]$$

Defining

$$A = \bar{U}T, \quad B = P\bar{V}, \quad G = T^{-1}\Lambda^{-1}P^{-1}, \qquad [5]$$

we obtain

$$\ddot{X} = AGB. \qquad [6]$$

The factor analytic solution is now a product of three matrices instead of the usual solution, $X = AB$. This seems to be a loss of parsimony, but this loss is compensated by a gain: both scales and concepts are treated symmetrically. We have a factor loading matrix for scales and a factor loading matrix for concepts. And both matrices have been transformed to some especially meaningful form such as simple structure.

In the formula AGB the two matrices A and B have a symmetrical role and both are analogous to factor loadings, that is, they can be explained as weights applied to factor scores. Therefore, G functions as a factor score matrix. But, unlike the usual factor analytic formula for the factor scores, the matrix G is reduced with respect to both rows and columns. To return to the semantic differential example: both scales and concepts are reduced to factors, and the internal core matrix G gives the factor scores of concept factors on scale factors.

A test by subjects paradigm may clarify the issue in more familiar terms, though the requirement for symmetry is not as cogent here. It is also convenient to use in this context the terminology "idealized entities," like "idealized subjects," and "idealized traits," (Helm & Tucker, 1962, p. 439; Tucker & Messick, 1960, p. 5).

In the fundamental equation of classical factor analysis,

$$ {}_iX_k = {}_iA_m B_k, \qquad [7]$$

${}_iX_k$ is a matrix of subjects by tests, ${}_iA_m$ is a matrix of subjects by factors, and ${}_mB_k$ is a matrix of factors by tests. In the matrix ${}_mB_k$, each row can be interpreted as the entries of an "idealized subject." Employing this terminology, each idealized subject is a k-tuple consisting of the loadings of the k tests on a factor.

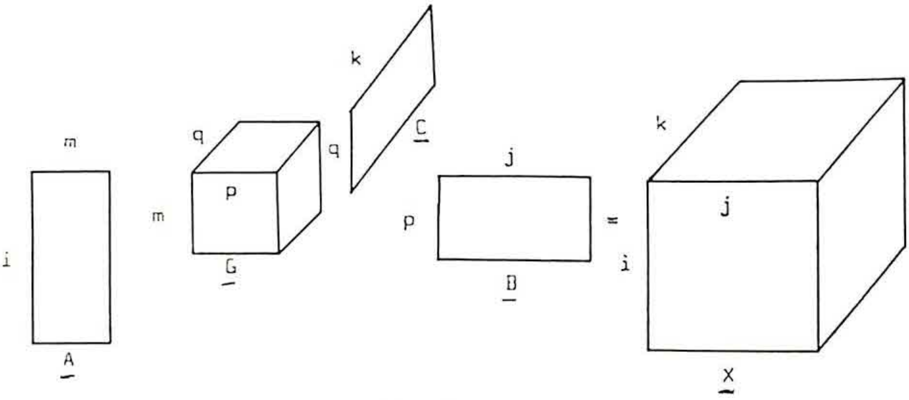


Fig. 1

Sectioning matrix B by rows we obtain:

$$\begin{aligned}(x_{j1}, x_{j2} \cdots x_{jk}) = a_{j1}\,&(b_{11}, b_{12} \cdots b_{1k})\\ &+ a_{j2}(b_{21}, b_{22} \cdots b_{2k})\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ &+ a_{jm}(b_{m1}, b_{m2} \cdots b_{mk}).\end{aligned} \qquad [8]$$

The row $(x)_j$ is a vector of predicted scores for a given subject j on the k tests. This row-vector is a weighted sum of the rows of B. The rows of B are the scores of the set of idealized subjects, and the row vector $(a)_j$ is a row of weights for subject j.

Now we have to modify this interpretation in order to adapt it to two-mode factor analysis with an inner core matrix. In the formula

$${}_iX_k = {}_iA_m G_m B_k \qquad [9]$$

the roles of subjects and tests are symmetric. ${}_mG_m$ is a matrix of scores of idealized subjects on idealized tests. The rows of G are for idealized subjects and the columns are for idealized tests. The A and B matrices are weights. Each row of the matrix ${}_iA_m$ is a set of weights for a given subject. Applying these weights to the idealized subjects (i.e., to the rows of G) we obtain the scores of the subject expressed as a linear combination of the idealized subjects. Similarly, each column of ${}_mB_k$ is a set of weights for a given test. Applying these weights to the idealized tests (i.e., the columns of G) we obtain the scores on the test expressed as a linear combination of the idealized tests. Obviously, the pre- and postmultiplication of G by A and B produces the scores (or predicted scores if nonsignificant factors are omitted) of all subjects on all tests from the scores of the idealized subjects on the idealized tests.

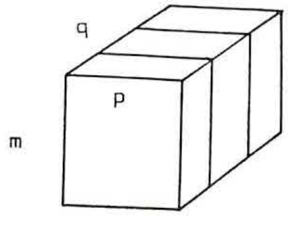


Fig. 2

An Extension to Three-Mode Factor Analysis

We can extend this formulation to three-mode factor analysis. The fundamental formula of three-mode factor analysis is:

$$x_{ijk} = \sum_m \sum_p \sum_q a_{im} b_{jp} c_{kq} g_{mpq}. \qquad [10]$$

The three-mode factor analysis starts with data given in a three-way classification. These data are reduced to an inner core G_{mpq}, which is again of three-way classification, and three factor weight matrices, ${}_iA_m$, ${}_jB_p$, ${}_kC_q$. The inner core, G, is analogous to the inner core in the equation $X = AGB$. The A, B, and C matrices are analogous to the factor loading matrices. These relations are described diagrammatically in terms of boxes and rectangles in Figure 1. The box G can be sectioned into q frontal matrices of order $m \times p$ each, as in Figure 2. Each of these is premultiplied by A and postmultiplied by B. The resulting box can now be sectioned laterally into j matrices of order $i \times q$ each. Each of these can now be postmultiplied by C as shown in Figure 3.

These operations reproduce the box X of order $i \times j \times k$. The multiplication of G_{mpq} by the three linear operators A, B, and C, can be performed in any order; that is, we can start with multiplication of G by A, then by B, and finally by C. We can also start with C, then multiply by B, and finally by A, and so on. This follows from the commutativity of multiplication and of summation in the fundamental formula, 10.

Now the generalization of two-mode factor analysis becomes obvious. Let us return to the example of the semantic differential. The data are given as concepts by scales by subjects (raters). We obtain three weight matrices (or factor loading matrices): one for concepts, one for scales, and one for subjects. This establishes three sets of factors. Then, we get an inner core "box." In analogy to the two-mode case, this gives the interrelations that connect the three sets of factors, that is, the factor score of each concept factor on each scale factor for each subject factor.

Employing the terminology of idealized entities, the inner core box gives the scores of each idealized concept on each idealized scale for each idealized subject; and the three outer matrices are weights applied to these idealized entities needed to reproduce the original scores of subject per scale per concept.

The Semantic Differential

Data

This study is a reanalysis of semantic differential data collected by Edward Ware (1958). There are 31 concepts rated on 20 scales by 60 subjects. The subjects are high-school students sampled according to the scheme indicated in Table 1.

The scales had seven categories (1–7) and were presented in graphical format.

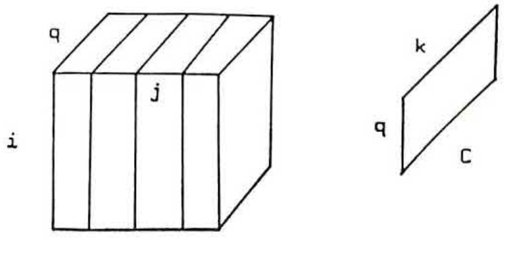


Fig. 3

TABLE 1

DISTRIBUTION OF SUBJECTS IN THE SEMANTIC DIFFERENTIAL STUDY BY SEX AND IQ

	Male	Female	Totals
High IQ	15	15	30
Low IQ	15	15	30
Total	30	30	60

Method

Scoring. The data were rescored from a 1 to 7 scale to a +3 to −3 scale. Assuming that 0 (the scale midpoint) is a natural neutrality point, and assuming that differences in size of unit (or standard deviation) reflect individual differences that ought to be retained in the analysis, no additional standardization seems to be required. The scores were, therefore, treated as standard scores, and sums of products were employed throughout, instead of correlations.

There are two ways to convert the data to standard scores that can be suggested as intuitively meaningful. One method is standardization of each scale as a variable, pooled over concepts and subjects. The other one is a separate standardization of each subject, which is essentially a conversion to ipsative scores. The effects of the mean and the unit of measurement have to be discussed separately.

The mathematical formulas expressing the effects of elimination of the mean (or any additive constant) are quite cumbersome. Fortunately, the deviations of the mean of each scale from zero were negligible so that this issue did not have practical significance.

On the other hand, the effect of the unit of measurement is quite simple. It follows from Formula 10 that a multiplicative constant can be incorporated in any of the matrices A, B, C depending on its subscripts (i.e., on the method of standardization). As shown in the next section, the score matrices of each subject are proportional; and by a conversion of each subject to a standard scale, we would lose some interesting information.

Factor Loading Matrices. Let i be the subscript for subjects, j for concepts, k for scales; i', j', and k' are used as alternative subscripts to i, j, and k. Three matrices of sums of products were computed:

$$\sum_j \sum_k x_{ijk} x_{i'jk}, \qquad [11]$$

$$\sum_i \sum_k x_{ijk} x_{ij'k}, \qquad [12]$$

$$\sum_i \sum_j x_{ijk} x_{ijk'}. \qquad [13]$$

These matrices were analysed by the method of principal axes and rotated to simple structure. The resulting factor loading matrices are the matrices A, B, and C of Equation 10. The inner core is obtained as a least square solution[4] (Levin, 1963).

---

[4] L. R Tucker, Analysis procedures in three-mode factor analysis, in preparation.

TABLE 2

FOUR SCALE FACTORS—VARIMAX ROTATION

Scales	Factor I	Factor II	Factor III	Factor IV
1. Fast-Slow	−05	58[a]	10	15
2. Even-Uneven	06	06	67[a]	03
3. Strong-Weak	22	66[a]	00	−02
4. Good-Bad	61[a]	−01	46[a]	−08
5. New-Old	06	−12	24	10
6. Warm-Cool	05	02	04	54[a]
7. Moral-Immoral	52[a]	11	11	−01
8. True-False	56[a]	20	−03	−06
9. Predictable-Unpredictable	02	−05	61[a]	−16
10. Sociable-Unsociable	40[a]	12	43[a]	09
11. Hot-Cold	−06	04	03	52[a]
12. Hard-Soft	−41[a]	46[a]	−13	−26[a]
13. Fair-Unfair	60[a]	−09	−15	03
14. Usual-Unusual	66[a]	05	−13	−04
15. Kind-Cruel	56[a]	−17	22	07
16. Straight-Twisted	16	−05	28[a]	04
17. Powerful-Powerless	24	55[a]	−23	08
18. Colorful-Colorless	33[a]	09	19	19
19. Simple-Complex	16	−25	05	−23
20. Emotional-Unemotional	14	28[a]	−32[a]	38[a]

[a] Salient loadings.

Results

Number of Factors. The number of factors has been determined, primarily, on basis of the latent roots. If we plot the latent roots in decreasing order of magnitude we find that the roots drop abruptly at the beginning, then taper off slowly. The bending point is used as a cutting point to determine the number of factors.

There is, apparently, only one major factor for subjects. The latent roots drop abruptly after the first root: 45,900; 5,500; 4,460; 3,650; 3,470; 3,270; 3,110; 2,290; etc. Correspondingly, only one idealized subject is retained in the inner core. Consequently, for these data, the predicted scores of each subject are computed by multiplication of the idealized subject by a constant of proportionality. Employing a somewhat different procedure, Edward E. Ware arrived at the same conclusion.[5]

Four factors have been retained for concepts. The latent roots are 26,890; 22,520; 9,470; 7,440; 4,810; 4,340; 3,870; 3,800; etc. From the fourth root onward the roots taper off gradually and slowly.

The results for scales are not so clear-cut. The latent roots are 31,090; 18,140; 10,910; 8,920; 7,670; 6,510; 5,990; 5,760; etc.

There are either four or five factors. Considerations of meaningful interpretation support the decision to retain four factors for scales. The pattern obtained from a rotation of four scale factors provides a meaningful interpretation, and the factors can be identified in terms of factors commonly obtained in semantic differential studies. The pattern obtained by rotation of five scale factors was ambiguous.

Corresponding to the number of factors

TABLE 3

FOUR CONCEPT FACTORS—VARIMAX ROTATION

Concepts	Factor I	Factor II	Factor III	Factor IV
1. Food	38[a]	−07	12	13
2. Baby	52[a]	−08	15	−20
3. Pig	07	21	14	04
4. Grief	01	43[a]	00	−02
5. Silk	21	−12	30[a]	09
6. Snow	21	02	43[a]	01
7. Bottom	−02	12	14	18
8. Sickness	−02	50[a]	04	−12
9. Mother	64[a]	−14	03	17
10. Army	20	15	−05	45[a]
11. Job	08	22	21	−23
12. Jelly	15	−10	43[a]	20
13. Street	14	10	17	35[a]
14. Fear	03	53[a]	01	13
15. Anger	04	54[a]	−15	14
16. Insane man	−08	50[a]	−06	−04
17. Me	45[a]	−05	00	15
18. Cop	22	13	−04	36[a]
19. Sin	−10	55[a]	−09	05
20. Sleep	40[a]	−07	25	15
21. Pain	04	49[a]	−06	09
22. Spider	−07	33[a]	16	05
23. Butter	11	−12	44[a]	10
24. Lamp	23	−06	15	33[a]
25. Dream	34[a]	15	12	02
26. Chair	18	−13	28	25
27. Snail	02	02	39[a]	06
28. Sex	50[a]	15	01	13
29. Dirt	−02	14	36[a]	10
30. Statue	00	−04	22	39[a]
31. Beggar	−04	08	08	00

[a] Salient loadings.

---

[5] Personal communication.

retained for subjects, concepts, and scales, the inner core is of order 1 × 4 × 4.

Factor Loading Matrices and Interpretation of Factors. The factor matrices are given in Tables 2 and 3. Only the results of Varimax rotation (Harman, 1960, ch. 14) are reported here. The results of an oblique rotation were essentially the same.

The factors can be identified on the basis of salient variables. The scale factors are essentially similar to the major factors regularly obtained in semantic differential studies.

The first factor has high loadings on the following scales:

14. Usual-Unusual	66
4. Good-Bad	61
13. Fair-Unfair	60
8. True-False	56
15. Kind-Cruel	56
7. Moral-Immoral	52

This is the Evaluation factor.

The second factor has high loadings on the following scales:

3. Strong-Weak	66
1. Fast-Slow	58
17. Powerful-Powerless	55
12. Hard-Soft	46

This is a coalescence of the Potency and Activity factors. It was also obtained in a study by Triandis and Osgood (1958) and was called a Dynamism factor.

The third factor has high loadings on these scales:

2. Even-Uneven	67
9. Predictable-Unpredictable	61
4. Good-Bad	46
10. Sociable-Unsociable	43
20. Emotional-Unemotional	−32

and possibly,

16. Straight-Twisted	28

This factor seems to come closest to a Stability factor.

The fourth factor has few high loadings. It is high on these scales:

6. Warm-Cool	54
11. Hot-Cold	52
20. Emotional-Unemotional	38

Scales 6, 11, and 20 are virtually synonymous, especially when one considers their frequent metaphorical use. This factor may be called a Warmth factor. While the identification of this factor seems logically reasonable on basis of the data obtained here, it has not appeared as a separate factor in other studies of the semantic differential.

Since factor analytic studies of the semantic differential report scale factors only, the identification and interpretation of concept factors is based entirely on the results obtained in this study.

The first factor has high loadings on the following concepts:

9. Mother	64
2. Baby	52
28. Sex	50
17. Me	45

and possibly,

20. Sleep	40
1. Food	38
25. Dream	34

These concepts are apparently related to family and home surroundings and may be called a Human factor or perhaps an Intimacy factor.

The second factor is high on these concepts:

19. Sin	55
15. Anger	54
14. Fear	53
8. Sickness	50
16. Insane man	50
21. Pain	49
4. Grief	43

These concepts signify intense emotional states related to agitated behavior. This may be called an Agitation factor.

The third factor has high loadings on the following concepts:

23. Butter	44
12. Jelly	43
6. Snow	43
27. Snail	39
29. Dirt	36

and possibly,

5. Silk	30

TABLE 4

Core Matrix Corresponding to the Varimax Rotation

Concept factors	Scale Factors Evaluation I	 Dynamism II	 Stability III	 Warmth IV
Intimacy I	18	03	−01	18
Agitation II	−03	10	−19	03
Slickness III	12	−13	−06	−15
Toughness IV	03	21	09	−14

These concepts signify soft, compressible and slick matter of shapeless and malleable material. This factor may be called a Slickness factor.

The fourth factor has high loadings on these concepts:

10. Army	45
30. Statue	39
18. Cop	36
13. Street	35
24. Lamp	33

and possibly,

2. Baby	−20

The highest loadings on the fourth factor are somewhat lower than those of the three preceding factors. It seems that they have in common a property of Toughness, whether tactile, such as in Statue, or metaphorical, such as in Army. Note also the negative loading of Baby.

The Core Matrix. The core matrix is given in Table 4.

In order to facilitate interpretation, a categorization based on rank order is adopted here, namely, the lowest five entries, the highest five entries, and the middle six entries. In Table 5 this corresponds to the following cutting points:

High: |15| and above
Medium: between |03| and |15|
Low: |03| and below

TABLE 5

Core Matrix Corresponding to Varimax Rotation (Schematic Representation)

Concept Factors	Scale Factors Evaluation I	 Dynamism II	 Stability III	 Warmth IV
Intimacy I	++	0	0	++
Agitation II	0	+	−−	0
Slickness III	+	−	−	−−
Toughness IV	0	++	+	−−

The core matrices can be represented schematically employing the following notation:

++	for high positive
+	for medium positive
0	for low or neutral
−	for medium negative
−−	for high negative

The schematic representation is given in Table 5.

The scores of the Agitation factor require a few comments. A similar relation between the concept "Mental disorder" and stability scales is reported by Nunally (1961). On the other hand, the neutral (or rather low negative) score of Agitation on the Evaluation factor defies explanation. An attempt to employ Oblimax rotation produced a shift downwards but not large enough to pull it out of the neutrality range.

The other entries of the table need no further comment.

The S-R Inventory of Anxiousness

Data

This study is a reanalysis of the data employed by Endler, Hunt, and Rosenstein (1962). This inventory consists of 14 responses to 11 situations. Each response is matched with each situation. The total number of items is 154.

> The choice of the situations . . . is based on an intuitive attempt to select a variety of situations that would be familiar through either direct or vicarious experience to most college freshmen or sophomores (the *S*s), a variety which would vary from the typically innocuous to the quite threatening.
>
> The S-R inventory employs a five-step scale, ranging from *none* to *very much*, with which *S* is asked to register the felt intensity of his own response to the situation in question [Endler et al., 1962, p. 5].

Here is an example of an item:

You are about to go on a roller coaster.
Heart beats faster 1 2 3 4 5
Not at all Much faster

The subjects were 169 male and female students in elementary courses in psychology at Pennsylvania State University.

Standardization of Scores. Unlike the semantic differential, it would be untenable to assume here that the midpoint of the scale of each item corresponds to the origin, neither can we assume that the units of measurement are equal.

The scores were converted to standard scores by the usual procedure of subtraction of the mean and division by the standard deviation. Note that there are $11 \times 14 = 154$ variables and there are 169 subjects for each variable.

Computation of Factor Loadings and Core. Though based on the same mathematical principles, a different method was employed here than with the semantic differential. The procedure provides another aspect of comparison between three-mode factor analysis and some classical procedures (Mahmoud, 1955).

Three correlation matrices were computed: (*a*) The intercorrelation of all items. This is a matrix of order 154. (*b*) The average intercorrelation of the situations. This is a matrix of order 11. (*c*) The average intercorrelation of the responses. This is a matrix of order 14. Each of these matrices was factor analyzed by by the method of principal axes.

The principal axes of the large matrix (154×154) were employed to compute the inner core.

The analysis of the 154 items (11 situations by 14 responses) yielded three significant factors. At this point, it seems that these factors generate a confounding of two logically distinct modes; for the loadings of each item involve both a situation and a response. However, rearranging the loadings of each factor in the format of a matrix of 11 situations by 14 responses, we obtain again a three-way classification table consisting of factors by situations by responses, thus keeping the distinct modes neatly separated. This is the first step to obtain the inner core. The factors correspond to the idealized subjects of our previous terminology. We still have two additional matrices: the principal axes of the situations and the principal axes of the responses. They can serve as linear operators to complete the computation of the inner core and to reduce the situations to idealized situations and the responses to idealized responses. The detailed procedure to obtain the inner core is discussed elsewhere (Levin, 1963; Tucker[4]).

The principal axes matrices of the situations and of the responses were rotated obliquely (Tucker, 1944; Tucker, 1955). The results obtained correspond to the matrices B and C of Equation 10. Due to methodological restrictions, the matrix A corresponding to subject factors could not be computed.

Results

Number of Factors. Three factors were retained for situations and three factors for responses.

The latent roots of the situation matrix are: 3.54, .66, .47, .32, .28, .24, .13, .10, .08, .08, .04. The roots taper off slowly after the third root. The sum of the first three roots is 4.67. The first three roots account for 78% of the variance.

The latent roots of the response matrix are: 3.86, 1.55, .69, .38, .34, .28, .12, .10, .05, .03, .00, −.03, −.05. The abrupt drop between the third and fourth root, and the slow tapering off from the fourth root onward is here more prominent. The sum of the first three roots is 6.10. The total sum of the roots is 7.56. The first three roots account for 81% of the variance.

The latent roots of ${}_{154}R_{154}$ are: 23.89, 7.39, 6.37, 4.76, 4.28, 4.03, 3.50, 3.24, 2.95, 2.83, etc. There is a sharp drop between the first and second root, but there is also a relatively large decrease from the third to the fourth root. From the fourth root onward there is a slow gradual decline. This can be seen clearly from the differences between each pair of successive roots: 16.50, 1.02, 1.61, .48, .25, .53, .26, .29, .12. The order of magnitude of the first three differences is larger than that of the remaining differences. This points toward the conclusion to retain three factors. Correspondingly, the inner core is, thus, a $3 \times 3 \times 3$ box.

TABLE 6

Factor Loadings of Situations—Oblique Rotation

	I	II	III
1. Auto trip	.01	.04	.38[a]
2. New date	.31[a]	−.12	.30[a]
3. Psychological experiment	.00	.03	.54[a]
4. Ledge high on mountain side	−.06	.56[a]	.01
5. Speech before large group	.55[a]	−.10	.03
6. Counseling Bureau for personal problems	.20	.04	.36[a]
7. Sail boat on rough sea	−.01	.42[a]	.08
8. Competitive contest	.46[a]	.09	−.14
9. Alone in woods at night	.07	.47[a]	−.06
10. Interview for important job	.52[a]	−.03	.04
11. Final exam in important course	.33[a]	.09	.05

[a] Salient loadings.

The Factor Loading Matrices. The rotated factor loadings are given in Tables 6 and 7.

The salient variables of the first factor for situations are:

5. Speech before large group	.55
10. Interview for important job	.52
8. Competitive contest	.46
11. Final exam in important course	.33
2. New date	.31

These are situations where interpersonal stress relations are involved. This factor was called Interpersonal.

TABLE 7

Factor Loadings of Responses—Oblique Rotation

	I	II	III
1. Heart beats faster	.63[a]	.10	−.04
2. Get "uneasy feeling"	.68[a]	−.02	−.05
3. Emotions disrupt actions	.46[a]	−.04	.10
4. Feel exhilarated and thrilled	.11	.60[a]	−.02
5. Want to avoid situation	.34[a]	−.38[a]	.03
6. Perspire	.34[a]	−.03	.18
7. Need to urinate frequently	.07	.09	.49[a]
8. Enjoy the challenge	−.04	.77[a]	.00
9. Mouth gets dry	.21	−.03	.31[a]
10. Become immobilized	.23	−.05	.25
11. Get full feeling in stomach	.05	.05	.42[a]
12. Seek experiences like this	−.05	.73[a]	.05
13. Have loose bowels	−.15	.01	.55[a]
14. Experience nausea	.02	−.09	.41[a]

[a] Salient loadings.

The salient variables of the second factor for situations are:

4. Ledge high on mountain side	.56
9. Alone in woods at night	.47
7. Sail boat on rough sea	.42

These are situations of concern with inanimate personal danger. This factor was called Inanimate.

The salient variables for the third factor for situations are:

3. Psychological experiment	.54
1. Auto trip	.38
6. Counseling Bureau for personal problems	.36
2. New date	.30

The similarity of "Psychological experiment" and "Counseling Bureau" is obvious. A "New date" has a component of similarity to "interpersonal stress." Yet its affinity to the anticipation of a psychological experiment seems cogent. These items may, perhaps, be characterized as comprising situations of "intrapersonal stress." The meaning of "Auto trip" in this context is not clear. For brevity this factor may be labeled Psychological.

The salient variables of the first factor for responses are:

2. Get "uneasy feeling"	.68
1. Heart beats faster	.63
3. Emotions disrupt actions	.46
6. Perspire	.34
5. Want to avoid situation	.34

This factor was called a Distress factor.

The salient variables of the second factor for responses are:

8. Enjoy the challenge	.77
12. Seek experiences like this	.73
4. Feel exhilarated and thrilled	.60
5. Want to avoid situation	−.38

This factor was called an Exhilaration factor.

The salient variables of the third factor for responses are:

13. Have loose bowels	.55
7. Need to urinate frequently	.49
11. Get full feeling in stomach	.42
14. Experience nausea	.41
9. Mouth gets dry	.31

This factor was called Autonomic.

The Core Matrix. The core is a three-mode matrix of order 3 × 3 × 3, that is, it consists of three idealized subjects or types. Each type is an interaction matrix of Situation Factors × Response Factors, that is, for each type we obtain the scores of response factors to the situation factors. The core matrix is given in Table 8. It is arranged in sections corresponding to types. Each section is a 3 × 3 matrix with rows corresponding to response factors and columns to situation factors. Since the items were scored on a scale from "not at all" to "very much"; a low negative entry corresponds to "not at all," a high positive entry corresponds to "very much," and a zero (or near zero) entry corresponds to the average (the scores were converted to standard scores at the first step of the computation).

TABLE 8

CORE MATRIX—OBLIQUE ROTATION

	Interpersonal	Inanimate	Psychological
Type I			
Distress	.40	.39	.22
Exhilaration	−.05	−.11	.03
Autonomic	.41	.37	.23
Type II			
Distress	−.07	−.29	.20
Exhilaration	.29	.44	.32
Autonomic	.19	.05	.33
Type III			
Distress	−.22	.30	−.16
Exhilaration	.37	−.19	.19
Autonomic	−.01	.37	−.15

The first type is high in Distress response (.40, .39, .22) and Autonomic response (.41, .37, .23), and low to low-average in Exhilaration (−.05, −.11, .03). Comparing situations, we find a stronger reaction to the Interpersonal and Inanimate situations than to the Psychological situation.

The second type is notably high in Exhilaration (.29, .44, .32), which is especially conspicuous in the Inanimate situation; but the Distress and Autonomic responses depend on the situation. The anxiousness responses to the Interpersonal factor are somewhere between the Inanimate factor and the Psychological factor: average Distress (−.07) and somewhat high Autonomic (.19). The Exhilaration response (.29) is almost the same as to the Psychological situation (.32).

Comparing situations, we find a very high Exhilaration response (.44) and a low Distress response (−.29) to the Inanimate situation. The Autonomic response is average (.05), but it is not necessarily an anxiousness response—it might be a seasickness type of reaction (note Situations 4 and 7). The response to the Psychological situation is high in all factors: Exhilaration (.32) and the anxiousness factors, that is, Distress and Autonomic (.20, .33). This conforms to the hypothesis in the monograph by Endler et al. (p. 5) that excitement may come both in positive and negative forms.

The third type has high Distress and Autonomic reactions (.30, .37) and low Exhilaration reaction (−.19) to the Inanimate situations, and conversely high Exhilaration (.37, .19) and low to low-average Distress and Autonomic reactions to the Interpersonal and Psychological situations (−.22, −.01, −.16, −.15).

In a schematic and somewhat simplified summary the three types can be described as follows: The first type is noted for general anxiousness reactions: Distress and Autonomic reactions, though he is slightly less sensitive to the Psychological situation. The second type shows a thrilling reaction, especially to physical adventure, and a tendency to get excited to the other situations, that is, to react both with some exhilaration and anxiousness. The third type shows an anxiousness reaction to physical adventure and a thrilling reaction to the other situations.

CONCLUSION

The main contribution of three-mode factor analysis to classical factor analysis consists of the idea of an inner core and of the method to calculate its numerical values. The emperical studies reported here demonstrate that these values conform to a meaningful pattern. The inner core can be interpreted in terms of psychological types. Each type consists of a matrix which can be explained in terms of interaction among the factors pertaining to the other two modes. The semantic differential study demonstrates a special case consisting of one type only.

REFERENCES

Abelson, R. P. Scales derived by considerations of variance components in multi-way tables. In H. Gulliksen & S. J. Messick, (Eds.), *Psychological scaling: Theory and applications*. New York: Wiley, 1960. Pp. 169–186.

Burt, C. Test reliabilities estimated by analysis of variance. *British Journal of Statistical Psychology*, 1955, **8**(2), 103–118.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multi-trait-multimethod matrix. *Psychological Bulletin*, 1959, **56**, 81–105.

Cattell, R. B. The three basic factor-analytic research designs—Their interrelations and derivatives. *Psychological Bulletin*, 1952, **49**, 499–500.

Eckert, C., & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, **1**, 211–218.

Endler, N. S., Hunt, J. McV., & Rosenstein, A. J. An S-R inventory of anxiousness. *Psychological Monographs*, 1962, **76**(17, Whole No. 536).

Guttman, L. What lies ahead of factor analysis. *Educational and Psychological Measurement*, 1958, **18**, 497–515.

Harman, H. H. Analytical methods for the multiple-factor solution: Orthogonal case. In, *Modern factor analysis*. Chicago: Univer. Chicago Press, 1960.

Helm, L. E., & Tucker, L. R. Individual differences in the structure of color-perception. *American Journal of Psychology*, 1962, **75**, 437–444.

Levin, J. Three-mode factor analy[illegible]. Unpublishe doctoral dissertation, University of [illegible]nois, 1963.

Mahmoud, A. F. Test reliability in terms of facto theory. *British Journal of Statistical Psychology* 1955, **8**(2), 119–135.

Nunnally, J. *Popular conception of mental health*. Nev York: Holt, Rinehart & Winston, 1961.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: Univer. Illinoi Press, 1957.

Triandis, H. C., & Osgood, C. E. A comparative factorial analysis of semantic structures in monolingua Greek and American college students. *Journal o Abnormal and Social Psychology*, 1958, **57**, 187–196

Tucker, L. R. A semi-analytical method of factorial rotation to simple structure. *Psychometrika*, 1944 **9**, 43–68.

Tucker, L. R. The objective definition of simple structure in linear factor analysis. *Psychometrika*, 1955 **20**, 209–225.

Tucker, L. R. The extension of factor analysis t three-dimensional matrices. In N. Frederiksen (Ed.), *Contributions to mathematical psychology* New York: Holt, Rinehart & Winston, 1964.

Tucker, L. R, & Messick, S. Individual differences i multidimensional scaling. Research Memorandum No. RM-60-15, 1960, Educational Testing Service Princeton, New Jersey.

Ware, E. E. Relationships of intelligence and sex to diversity of individual semantic spaces. Unpublished doctoral dissertation, University of Illinois, 1958.

(Received July 19, 1965)

17 MAR 1972